UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Functional aspects of hybrid genome interaction: sex determination and gene expression regulation in the *Squalius alburnoides* complex

Irene Alexandra de Carvalho Braz dos Santos Pala

DOUTORAMENTO EM BIOLOGIA
(Biologia Evolutiva)

2008

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Functional aspects of hybrid genome interaction: sex determination and gene expression regulation in the *Squalius alburnoides* complex

Irene Alexandra de Carvalho Braz dos Santos Pala

Tese orientada por:

Professora Doutora Maria Manuela Coelho
Professor Dr. Manfred Schartl

DOUTORAMENTO EM BIOLOGIA
(Biologia Evolutiva)

2008

The work presented in this dissertation was developed with the support of the Fundação para a Ciência e a Tecnologia: PhD Fellowship (SFRH/BD/19286/2004) and by the Projects POCTI/BSE/40868 (FEDER participation) and PTDC/BIA-BDE/ 65586/2006.

Nota prévia

Para a elaboração da presente dissertação, e nos termos do nº 1 do Artigo 40, do regulamento de Estudos Pós-Graduados da Universidade de Lisboa, publicado no *Diário da República* nº 153, II Série de 5 de Julho de 2003, foram usados integralmente artigos científicos publicados (2) ou a submeter para publicação (2) em revistas internacionais indexadas. Tendo os trabalhos referidos sido efectuados em colaboração, a autora da dissertação esclarece que participou integralmente na concepção e execução do trabalho experimental, na análise e interpretação dos resultados e na redacção de todos os manuscritos.

À minha família

Contents

AGRADECIMENTOS

Durante este projecto de quatro anos foram muitas as pessoas que a nível profissional e pessoal marcaram determinantemente este percurso e às quais gostaria de deixar o meu sincero agradecimento:

À Professora Doutora Maria Manuela Coelho pelo apoio e pela confiança que sempre depositou em mim, pela liberdade que me permitiu no desenvolvimento deste projecto e pela motivação e apoio que sempre me transmitiu nos momentos de maior dificuldade.

To Professor Manfred Schartl, for embracing this project, for opening the doors of his lab to make it possible and for his cooperation, guidance, availability and support, at all times. It was a pleasure and a great learning experience to work with you.

À Professora Doutora Maria João Collares-Pereira pelas sugestões e pelo constante interesse na evolução deste trabalho.

À Professora Doutora Sólveig Thorsteinsdóttir pela participação, aconselhamento e valiosa discussão de ideias em várias fases do desenvolvimento deste projecto.

Ao Doutor Miguel Brito e à Joana Vacas pelo aconselhamento e permanente disponibilidade para o auxílio na realização de experiências de Real-Time PCR.

À Doutora Maria Judite Alves, por ter sempre disponibilizado todos os meios para a realização de alguns pontos deste trabalho e pelo interesse sempre demonstrado.

À minha colega Carina Cunha, que partilhou comigo as "facilidades" de estudar o *alburnoides*. Pelas disussões, pela partilha do trabalho duro, dos dias longos. E pelas brincadeiras. E principalmente por ser muito boazinha.

À Marta Gromicho, outra vítima da hibridação e poliploidia, salvo seja, pela partilha de experiências, pelas discussões e pelo companheirismo criado ao longo destes anos.

À Cristina Luís, o olhar do outro lado da barricada. Pela ajuda em todas as ocasiões, pela boa disposição, pela presença e pelo incentivo. E por ser um público maravilhoso para as minhas parvoíces!

A todos os colegas com quem partilhei o laboratório durante estes anos, pela troca de experiências, pela boa disposição e pelo excelente ambiente de trabalho. Em particular aqueles que tiveram uma partilha mais directa do trabalho de todos os dias, dos animados almoços e de muitos jantares: Ana Rita Amaral, Ana Rita Grosso, Ana Gaspar, Anabel Perdices, Carina Cunha, Cristiane Silveira, Cristina Luís, Elisabete Malveiro, Fernanda Bajanca, (Prof B') Gabriela Rodrigues, Joana Abrantes, João Santos, João Nunes, Maria Ana Aboim, Maria João Pereira, Marta Gromicho, Mónica Silva, Natacha Mesquita, Patrícia Salgueiro, Pedro Campinho, Pedro Rifes, Romina Henriques.

Aos colegas que sempre me auxiliaram nas atribuladas saídas de campo, pela preciosa ajuda a compensar a minha inexperiência, pelos ensinamentos e disponibilidade: Anabel Perdices, Carina Cunha, Filipe Ribeiro, Luís Costa e Maria Ana Aboim.

À Dª Branca pela ajuda e disponibilidade e pela boa disposição. À Dª Irene por me ter ajudado em tudo o que esteve ao seu alcance.

To all my friends and colleagues at the Biozentrum in Würzburg, for their warm welcome, and for the advice, help and friendship during these four years. A special word to Nils Klüver for introducing me to the lab work, the cooperation and the company during these years.

À minha família e amigos, por compreenderem o tempo que não tive, os momentos em que não pude estar presente. E por terem tornado os momentos que tive com eles ainda mais preciosos.

À minha avó por todo o entusiasmo desde sempre demonstrado por tudo o que faço.

Aos meus pais, pelo apoio e afecto incondicionais em todos os momentos. Principalmente nos mais difíceis. E também nos mais felizes.

À minha irmã pelo entusiasmo, por acreditar sempre em mim e por compensar as minhas ausências durante este período tão atribulado. É tão bom termos alguém que pensa que somos uns grandes investigadores!!! É pena que não faças parte do júri…

Ao Zé, companheiro de todos os momentos. Por estares ao meu lado hoje, como sempre, nesta nossa nova conquista.

RESUMO

Os processos de hibridação e poliploidia, estão na base de numerosos fenómenos evolutivos. A questão de como se constitui e mantem um genoma poliplóide e de quais as perspectivas de evolução de um sistema de cópias génicas supranumerárias tem desde sempre despertado o maior interesse e fascínio. As consequências evolutivas dos processos de hibridação e poliploidia e a sua influência na evolução e especiação têm sido amplamente avaliadas em vários grupos de poliplóides. A análise dos efeitos destes processos na expressão génica e na regulação das cópias de cada genoma tem merecido especial atenção nas plantas mas, no que diz respeito aos vertebrados poliplóides, o impacto funcional do aumento do número de cópias genómicas é, na maior parte dos casos, desconhecido.

O complexo híbrido *Squalius alburnoides* apresenta características que o tornam particularmente interessante como base para o estudo dos efeitos da hibridação e poliploidia sobre a expressão de diferentes genomas reunidos num mesmo indivíduo. O complexo resulta da hibridação interespecífica e não recíproca entre fêmeas de *Squalius pyrenaicus* (genoma P) e de uma espécie próxima de *Anaecypris hispanica* (genoma A) e inclui formas diplóides e poliplóides cuja persistência e diversidade é mantida por diferentes modos de reprodução. É caracterizado por um acentuado desvio na distribuição dos sexos, com uma predominância de fêmeas triplóides na maior parte das populações e a ocorrência de uma linhagem aparentemente constituída exclusivamente por machos. Entre os organismos poliplóides, o complexo *S. alburnoides* é considerado um exemplo emblemático de sucesso, apresentando um sistema de trocas genéticas em que as diferentes formas diplóides e poliplóides participam activamente, contribuindo para a manutenção da diversidade e do potencial evolutivo das diferentes populações. No entanto, em *S. alburnoides* tal como em outros vertebrados poliplóides, o impacto dos processos de hibridação e de aumento do número de cópias genómicas na dinâmica de expressão dos genes originários de diferentes genomas raramente foi estudado, assim como a sua potencial influência em mecanismos aparentemente afectados pelos processos de hibridação, como o equilíbrio na distribuição dos sexos.

Com esta dissertação pretendeu-se obter os primeiros indícios dos efeitos funcionais da hibridação e da poliploidia sobre a regulação da expressão génica usando o complexo *S. alburnoides* como modelo de genoma poliplóide. A estratégia fundamental consistiu em dirigir essa análise para dois níveis distintos, essenciais para a funcionalidade e perpetuação dos indivíduos do complexo, e que no conjunto permitissem uma melhor perspectiva dos efeitos da hibridação sobre a regulação dos padrões de expressão de genes específicos. Numa primeira abordagem, foi dada especial atenção à base genética de um processo potencialmente conservado e estritamente regulado: a determinação

do sexo. Na base desta questão estão os acentuados desvios na distribuição dos sexos observados em *S. alburnoides* e a marcada associação de genótipos particulares a determinado sexo, cujo exemplo paradigmático é a ocorrência de uma linhagem de indivíduos de genótipo AA, aparentemente composta quase exclusivamente por machos. Numa perspectiva mais abrangente em termos dos padrões gerais de expressão génica ao nível do organismo poliplóide, pretendeu-se investigar a contribuição de cada genoma presente nos híbridos para os padrões globais de expressão de genes individuais.

Partindo destas duas ideias base, esta dissertação teve como objectivos: 1. A identificação e isolamento de genes que potencialmente poderiam estar envolvidos no processo de determinação do sexo em *S. alburnoides* e de genes com expressão generalizada no organismo, criando uma base molecular para o estudo dos níveis de análise inicialmente propostos; 2. A caracterização dos padrões de expressão dos genes potencialmente implicados no processo de determinação do sexo na gónada adulta e durante o desenvolvimento, em *S. alburnoides* e na espécie parental bissexuada *S. pyrenaicus*; 3. A comparação dos padrões de expressão de genes individuais em *S. alburnoides* e *S. pyrenaicus*, cuja presença em determinadas localizações espacio-temporais sugerisse a participação nos processos de determinação do sexo e/ou diferenciação e manutenção da integridade da gónada. O objectivo seria aferir a existência de diferenças capazes de justificar os desvios observados nos híbridos; 4. Caracterizar a contribuição de cada genoma heteromórfico presente nos híbridos de *S. alburnoides* com base na análise da expressão de alelos específicos de cada genoma, em genes com expressão generalizada (*housekeeping*) ou restrita a determinados tecidos; 5. Correlacionar os padrões de expressão com a história evolutiva do complexo e aprofundar o conhecimento acerca do impacto da regulação da expressão génica no contexto global da evolução dos poliplóides.

Relativamente à determinação do sexo, foi possível estabelecer uma base molecular para o estudo do processo, através do isolamento em *S. alburnoides* e *S. pyrenaicus*, de genes conservados nas cascatas de determinação do sexo de vários vertebrados: *vasa*, *amh*, *dmrt1*, *wt1*, *dax1* e *figla*. Pela primeira vez foi possível identificar e comparar a contribuição destes genes nas gónadas adultas de ambos os sexos de diferentes formas do complexo e da sua espécie parental *S. pyrenaicus*. O processo de hibridação e o aumento de ploidia aparentemente não afectaram os padrões de expressão destes genes, e a sua potencial contribuição em tipos celulares específicos foi mantida inalterada nos híbridos. Para a maior parte dos genes, os padrões de expressão são concordantes com descrições feitas noutros teleósteos, mas em alguns casos, nomeadamente para *dmrt1* e *figla,* foram descritos padrões que poderão implicar diferentes funcionalidades destes genes em *Squalius*.

Considerando a sua participação no processo de determinação do sexo, a expressão destes genes foi analisada durante o desenvolvimento de machos de *S. alburnoides*, e a presença de *amh*, *dmrt1* e *wt1* em fases iniciais do desenvolvimento e em localizações concordantes com o primórdio da gónada, sugerem uma participação destes genes na fase inicial de constituição destas estruturas. No contexto geral dos teleósteos, em que o conhecimento da regulação genética da determinação do sexo é na grande maioria dos casos escasso, a presente dissertação contribuiu com uma das raras abordagens a este problema num sistema poliplóide.

A análise mais abrangente da contribuição de cada genoma heteromórfico para os padrões globais de expressão de seis genes, revelou a ocorrência de regulação de expressão por silenciamento alélico, aqui descrita pela primeira vez, num vertebrado triplóide. Os padrões de expressão observados em indivíduos do Sul da distribuição de *S. alburnoides*, variáveis de acordo com o gene e o órgão analisados, não são concordantes com a hipótese de silenciamento de todo um genoma haplóide, mas sugerem um mecanismo mais plástico e menos estrito de regulação da expressão génica. Embora não tenha sido possível determinar o mecanismo exacto que está na base dos padrões observados, o silenciamento destes genes em *S. alburnoides* parece estar relacionado com um processo de compensação de dosagem, sugerido pela redução dos níveis dos diferentes transcritos para a dosagem diplóide, nos triplóides analisados. O impacto dos resultados obtidos não se limita ao complexo em particular, mas dá suporte a uma proposta mais abrangente no estudo da evolução dos poliplóides: a necessidade de manutenção de um equilíbrio quantitativo dos produtos de genes duplicados por poliploidização. No contexto global dos vertebrados, as evidências obtidas em *S. alburnoides* acrescentam ainda a possibilidade de que a capacidade de regulação efectiva da expressão génica no caso de aumento de ploidia, possa ser um dos factores que contribuem para a maior incidência de poliploidia entre os vertebrados inferiores e o sucesso na manutenção destas linhagens ao longo da evolução, em oposição a grupos como os mamíferos em que estas variações têm, na maior parte dos casos, efeitos deletérios.

A extensão desta análise a outras populações do complexo gerou resultados surpreendentes, sugerindo que os padrões de regulação génica podem variar dentro do complexo, de acordo com a origem geográfica e o tipo de genoma envolvido no processo de hibridação. Nas populações do Sul, em que os híbridos apresentam diferentes combinações dos genomas P (de *S. pyrenaicus*) e A (do ancestral paterno) foi observada uma tendência para a expressão preferencial de alelos específicos do genoma A. Pelo contrário, nas populações do Norte *de S. alburnoid*es (em que os genomas presentes são C, de *S. carolitertii* e A, do ancestral paterno), e para os mesmos genes, as diferentes formas diplóides, triploides e tetraplóides apresentaram, na maior parte dos casos, expressão bialélica dos genomas heteromórficos. O tipo de expressão génica nos híbridos das diferentes

populações foi comparado com a origem e a história evolutiva de cada população. Foi ainda analisada a potencial contribuição dos padrões de expressão como causa ou consequência do desempenho de cada linhagem na corrida para a adaptabilidade e especiação. Os resultados obtidos sugerem que vários percursos são possíveis na evolução dos organismos poliplóides, e que em contextos distintos, diferentes soluções podem ser adoptadas, no sentido da manutenção da variabilidade e da perpetuação de linhagens. Se nas populações bem adaptadas do Sul da distribuição de *S. alburnoides* a dinâmica de silenciamento de cópias alélicas parece contribuir para a manutenção de uma "relação" equilibrada entre os genomas P e A, nas diferentes populações do Norte os percursos evolutivos poderão seguir trajectórias distintas, e o desequilíbrio gerado pela presença de cópias genómicas ímpares poderá ser superado pela evolução por tetraploidização.

Na globalidade, os resultados obtidos nesta dissertação são um contributo significativo para o entendimento do papel da expressão génica e da co-regulação de genomas distintos reunidos pelos processos de hibridação e poliploidia para a evolução dos organismos poliplóides. As evidências recolhidas têm um óbvio impacto no conhecimento mais profundo de aspectos ainda não elucidados do sistema emblemático *S. alburnoides*. As constrições evolutivas, os desvios em processos conservados como a determinação do sexo, ou a necessidade de regular hierarquias genómicas que a partir de um evento particular se vêem confrontadas num mesmo núcleo não são exclusivas deste complexo e, como tal, as suas implicações estendem-se ao contexto mais amplo da funcionalidade de qualquer vertebrado poliplóide. Espera-se assim que esta dissertação constitua um ponto de partida para a exploração de um conjunto de questões evolutivas relevantes. No caso particular de *S. alburnoide*s, e uma vez criada uma base molecular para a exploração do processo de determinação do sexo, seria desejável identificar genes iniciais na cascata que possam estar na base dos desvios observados nos híbridos. Tendo por base a exploração e descrição obtidas durante este trabalho, espera-se ainda que seja possível investigar o mecanismo responsável pelos padrões de silenciamento observados. Numa perspectiva mais abrangente, seria também importante determinar se os fenómenos observados em *S. alburnoides* contribuem globalmente e de forma crucial, para a evolução de outros grupos de vertebrados poliplóides.

Palavras-Chave:

Hibridação; Poliploidia; Complexo *Squalius alburnoides*; Determinação do sexo; Expressão génica

ABSTRACT

Hybrid polyploidy success has been extensively reported in plant species and more restrictedly in animals, but the regulatory changes that contribute to genome stabilization and regulation in the presence of distinct chromosome sets are still elusive. Due to its unique features among polyploid taxa, the *Squalius alburnoides* complex of hybrid fish is a very desirable system to address these questions. It is considered an example of polyploid viability and evolutionary success, thus making it more interesting as a model to assess whether such features could be the result of an underlying specific gene regulatory basis. The aim of this dissertation was to assess whether hybridisation and polyploidy could have a functional effect over gene expression regulation: at the level of sex determination, a more strictly regulated process of gene interactions that apparently could be disrupted in the hybrids, as well as in a broader context regarding the contribution of each genome to the overall gene expression patterns.

Six conserved elements of sex determination cascades (*vasa*, *amh*, *dmrt1*, *wt1*, *dax1* e *figla*) were isolated, and their expression patterns and potential contributions were characterized in the adult gonads of *S. alburnoides* and the parental species *S. pyrenaicus*. Transcripts of four of these genes were shown to be present during early stages of *S. alburnoides* male development, suggesting their involvement, at least in the process of hybrid gonad differentiation.

The participation of genome-specific gene copies in the global gene expression patterns was assessed by the analysis of six housekeeping and tissue specific genes, revealing the occurrence of a mechanism of non-additive gene expression and allele silencing acting in *S. alburnoides* triploids, described for the first time in a polyploid vertebrate. Further examination of the evolutionary consequences of the observed patterns of gene expression and the extension of the analysis to representative populations of the complex's distribution, suggested that different evolutionary pathways might be followed by distinct lineages even within the same polyploid complex, and that gene expression might also vary accordingly. The present findings may constitute a starting point towards the elucidation of a complex evolutionary jigsaw: which factors might contribute to genome regulation and perpetuation in a polyploid species.

Keywords:

Hibridisation; Poliploidy; *Squalius alburnoides* complex; Sex determination; Gene expression.

CHAPTER 1 | Introduction

1. INTRODUCTION

The pathways of evolution are neither simple nor straightforward and hybridisation and polyploidy have always constituted a complex evolutionary jigsaw. If over the years several hypotheses accounting for the persistence and success of allopolyploids have been comprehensively addressed, much is yet to be uncovered regarding which would be the immediate functional consequences for individual genes and, globally, for genomes. In allopolyploid vertebrate systems the question of how genomes overcome the instability brought upon by the unorthodox combination of genome heterogeneity and chromosome number alteration has seldom been addressed. The present dissertation aimed at providing evidence of the impact of merging different genomes, with two main molecular based approaches applied to an emblematic allopolyploid – the Iberian *Squalius alburnoides* complex. The functional scars of the hybridisation and polyploidisation processes were investigated through an initial characterization of the basal and evolutionarily relevant process of sex determination in parallel with a more global depiction of gene expression patterns in the different forms that compose the complex. The evolutionary and functional implications of the patterns that have emerged from the analysis of gene behaviors in individuals of different genotype, ploidy, sex and geographical origin are not limited to *S. alburnoides*. A broader discussion is opened and potentially extended to other polyploids: how gene expression is being regulated in a system in which different genomes are combined at variable numbers and how this can dictate the lineage's evolutionary fate.

1.1. Polyploidy and hybridisation

One of the most dramatic sources of genome-scale variation is polyploidy, the heritable condition of possessing more than two complete sets of chromosomes. Polyploids can be divided in two main groups, depending on their chromosomal composition and the mechanism of formation: autopolyploids arise from chromosome doubling within an individual or from the merger of two similar genomes from different individuals within a species, while allopolyploids result from hybridisation between genetically distinct parents, usually different species (reviewed in Otto, 2007). Polyploidy is especially common among plants both as an ancient and an ongoing evolutionary process (Adams & Wendel, 2005). In animals, it has been proposed that the occurrence of polyploidy would be impeached by its interference with sex determination processes and the disruption of gene dosage imbalance (Orr, 1990). Nevertheless, a considerable number of stable polyploid animal

taxa have been revealed over the years in amphibians and reptiles (reviewed in Otto & Whitton, 2000), in fish (Leggat & Iwama, 2003) and even more rarely in mammals (Gallardo *et al.*, 2006). Polyploidy is particularly prevalent among hybrid taxa, in an association that is thought to result either from the high rate production of unreduced gametes by diploid hybrids (Ramsey & Schemske, 2002) or to function as a strategy to overcome the pairing problems between heterologous chromosomes (by providing each chromosomal set with a "suitable", compatible pair). Natural hybridisation has been shown to play a relevant role in both plant and animal evolution (Mallet, 2007). Particularly in fish, hybridisation is a quite widespread process, positively correlated with polyploidy (Le Comber & Smith, 2004) (Table I).

Table I. Summary of the occurrence of polyploidisation events in vertebrate taxa and respective modes of reproduction (when known). Adapted from Otto, 2007.

		Reproduction		
		Parthenogenesis	**Sexual**	**Unknown**
Groups	**Fish**	9	23	18
	Amphibia	3	16	1
	Reptiles	15	1	0
	Birds	0	0	0
	Mammals	0	1	0

The convergence of hybridisation with polyploidy often results in the establishment of asexual reproduction, and among vertebrates the first paradigmatic example of such system was described in the fish *Poecilia formosa* (Hubbs & Hubbs, 1932). In asexual systems, (sensu Beukeboom & Vrijenhoek 1998) reproduction is performed clonally, without sex and recombination (or with rare recombination) and might or not involve syngamy or karyogamy (Fig. 1a). These processes can include the production of unreduced eggs that can develop independently by clonal inheritance (parthenogenesis), or that have to be activated by sperm without syngamy (gynogenesis), and the transmission of only one particular set of chromosomes (hybridogenesis) (Dawley, 1989). Nevertheless, several hybrid and polyploid taxa have shown the capacity of incorporating mechanisms that allow for the introduction of novel genetic variability (Fig. 1b) such as random segregation and recombination between homospecific genomes (Günther *et al.* 1979; Alves *et al.* 1998; Arai & Mukaino 1998; Zhang *et al.* 1998; Kim & Lee 2000), designated as meiotic hybridogenesis (Alves *et al.*, 1998) and more rarely, low levels of recombination between parental genomes (Bogart 1989; Graf & Polls Pelaz 1989; Parker *et al.* 1989; Sites *et al.* 1990; Plötner & Klinkhardt 1992;Alves *et al.* 2002). Additionally, these mechanisms can involve the incorporation of

small parts of the paternal genome in the oocyte (microchromosomes), as evidenced in the gynogenetic *P. formosa* (Schartl *et al.*, 1995; Lamatsch *et al.*, 2004) and the incorporation of the total paternal genome, leading to the formation of polyploids (Alves *et al.*, 1998; Lamatsch, 2001).

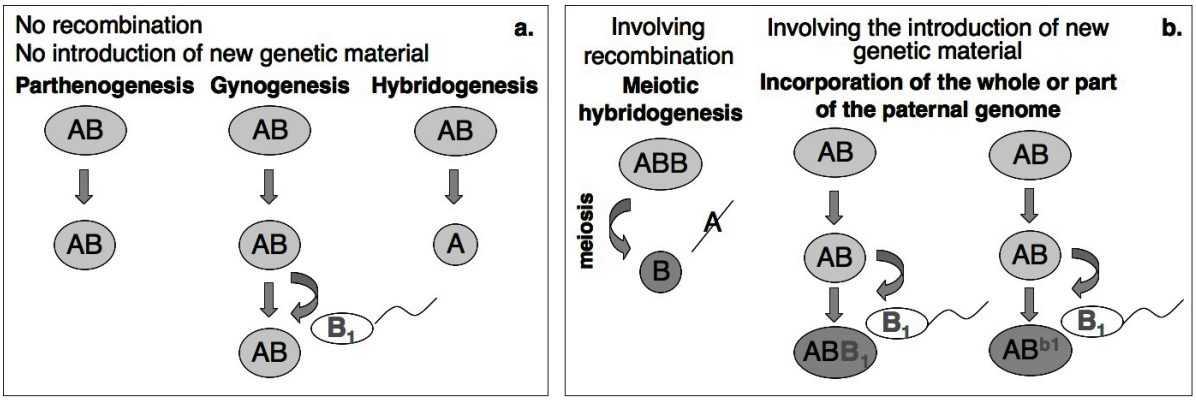


Figure 1. Summary of asexual reproductive modes. Reproduction and gamete formation processes were divided according to the presence or absence of recombination and the possibility of inclusion of novel genetic material.

The causes of evolutionary success of polyploid hybrid taxa have been the source of extensive debate, as classically they were thought to face important constrains in terms of variability maintenance and adaptation (Dawley, 1989). Ecologically based hypotheses have been postulated to account for such "unexpected" success, like the Frozen Niche-Variation Model (Vrijenhoek, 1979, 1984) in which clones are assumed to freeze the genetic variation of parental species. The recruitment of new clones through hybridisation would create a variation framework allowing specific niche adaptation and minimizing the overlap in the competition for resources. The probable causes seem however more complex, and the evidence provided from intensive study of the molecular basis of hybridisation and ploidy rise has yet opened new questions regarding the factors that could account for the persistence and evolution of polyploid taxa.

1.2. Polyploidisation- effects

Polyploidy has been regarded has an important force in plant evolution, by providing the basis for new phenotypical combinations and faster adaptative evolution (Adams & Wendel, 2005). In fact, evidence of polyploidy has been reported for numerous plant genomes (Blanc & Wolfe, 2004), and it is estimated that over 70% of angiosperms have undergone at least one round of genome duplication during the course of their evolution. One of the most prominent examples of ancient

polyploidisation events and their contribution to lineage evolution is the diploid *Arabidopsis thaliana*, in which a presently small genome has "covered up" cyclical genome doubling (Simillion *et al.*, 2002). In animals, the evolutionary impact of ploidy rise has also been addressed as an important question, despite the comparatively lower number of polyploid taxa (Otto & Whitton, 2000). In fact, it has been revealed that, as postulated by Ohno (1970), whole genome duplications would have occurred early in vertebrate evolution, (Panopoulou & Poustka, 2005; Dehal & Boore, 2005) constituting the basis for the speciation of teleost fish (Volff, 2005).

The common questions that arise from polyploid formation in every group, irrespective of taxonomic position, are how the new genomic arrangement is maintained and what should be the evolutionary fate of gene duplicates. Three outcomes are possible for a duplicate gene copy, of which only two allow the persistence of duplicates as functional entities: nonfunctionalization through deletion or degeneration, neofunctionalization (by which one member of the gene pair aquires a new function) and subfunctionalization (in which duplicates cooperate in the fulfillment of ancestral function) (Lynch & Conery, 2000). Most gene duplicates created by ancient polyploidisation were lost during evolution, but the retention and functionality of duplicate genes for millions of years (Nadeau & Sankoff, 1997) provides evidence for a possible role of polyploidy as a divergence source for evolutionary adaptation. The evidence of variable and biased process of gene retention has provided additional support for this idea, with some gene classes being more likely to be perpetuated than others (Hakes *et al.*, 2007; Chain *et al.*, 2008).

The evolutionary forces acting on several paleopolyploid taxa have apparently led to the return to a diploid context (despite the retention of some groups of duplicate genes), but the persistence of permanent polyploid lineages over long periods of time has motivated other evolutionary questions. The occurrence of stable vertebrate polyploids has also been shown to be more common than initially suspected, mostly in fish and amphibians (reviewed in Otto & Whitton, 2000) and several factors that might act as friends or foes in the evolution of polyploid lineages have been extensively investigated (Otto, 2007). The most "classically" reported effect of polyploidisation is the onset of asexual reproduction, impairing the possibility of recombination. Although presented as an argument for the mandatory "evolutionary dead-end" fate of these lineages (Dawley, 1989), asexual reproduction could also offer the advantage of reproduction in the absence of a sexual mate, and thus provide selective lead compared to parental species. Among other effects that have been accounted for as responsible for the evolutionary success of polyploid lineages, heterosis, the maintenance of a permanent heterozygote condition resulting from the absence of intergenomic recombination and gene redundancy have been pinpointed as most relevant. Gene redundancy not

only would offer the advantage of masking recessive alleles, but also bring about the additional possibility of diversifying gene functions (Comai, 2005).

Stepping away from the strict evolutionary consequences, yet other questions at a more functional level arise when addressing stable polyploids. Particularly in allopolyploidisation, the effects are not limited to simple merging of two distinct genetic hierarchies and often involve a wide spectrum of molecular and physiological arrangements, including changes in genome organization both at the genetic and epigenetic levels (Riddle & Birchler, 2003). These effects have been extensively characterized in plants and constitute the response to two main groups of factors: the gathering of two heterologous genomes and their regulatory frameworks within the same nucleus and the increased gene dosage when the process is accompanied by ploidy rise events. Genome reorganization events following allopolyploidisation include loss and gain of restricition fragment polymorphisms (RFLPs) (Song *et al.*, 1995; Liu *et al.*, 1998), the activation of transposable elements (Liu & Wendel, 2000) and gene conversion (Wendel *et al.*, 1995). Also, changes in epigenetic regulation of gene expression have been shown to occur immediately upon polyploidisation, including alteration of methylation patterns, gene silencing (Comai *et al.*, 2000; Lee & Chen, 2001; Kashkush *et al.*, 2002; Adams *et al.*, 2004) or non-additive contributions of genome-specific gene copies to overall expression (Adams *et al.*, 2003; Auger *et al.*, 2005).

The response to increased gene dosage has also been well characterized in plants and it has been proposed that a successful allopolyploid should have the ability to invoke and maintain diploid-like behavior (Ma & Gustafson, 2005). In a maize polyploid series it has been shown that most genes exhibited a dosage effect and several displayed dosage compensation. Non-linear relationships have also been observed in plant aneuploids[1], through unexpected gene expression deviations in haploid and triploid tissues, designated as odd ploidy response (Guo *et al.*, 1996), showing that the attempt to establish a new regulatory stable state in the polyploid genome is often a complex and heterogeneous process. In fact, silencing patterns can vary according to gene, and some genes have been shown to be silenced independently and repeatedly during polyploidisation, while others seem to follow stochastic processes (Adams, 2007). Thus, ploidy rise is frequently not accompanied by the expected increase in gene dosage, and the variability of non-additive gene expression profiles might be an important factor to account for gene perpetuation in polyploid genomes (Comai, 2005). These patterns relate to the recently re-discussed "gene balance" hypothesis, according to which a dosage-sensitive balance relationship is maintained in regulatory genes, ensuring the regulation of gene products to a "normal" dosage even in the presence of multiple gene copies. According to this hypothesis, not only the maintenance of dosage regulated gene expression patterns would allow

[1] Aneuploidy: the condition resulting from extra or missing chromosomes relative to the normal chromosome number of a species.

the retention of gene duplicates following polyploidisation, but also provide, by this means, a framework in which selection could differentially operate upon mutation (Birchler & Veitia, 2007). In plants, there has been an increasing interest in studying the various functional implications of hybridisation and polyploidy and their consequences at the evolutionary and adaptative levels, contrasting with a general lack of information on these aspects in polyploid animal taxa. In fact, and even though similar hybridisation and polyploidisation processes have also occurred, the gene regulation outcome and the patterns of genome interaction are still important open questions for allopolyploid vertebrates.

1.3. Sex determination-I: the multiple choices of teleost fish

1.3.1. Overview of sex determination

One of the most conserved biological phenomena is the establishment of the phenotypic sex of an individual, comprising two interrelated sets of mechanisms: sex determination and sex differentiation. The process of sex determination is controlled by genes that interact during development and regulate the cascade of events that promote the commitment of an undifferentiated gonad to a testis or ovary "fate", while the process of sex differentiation only starts subsequently, mediated by gonadal hormones (Nef & Parada, 2000).

Despite its conservation as a global process, sex determination can be modulated by a wide diversity of factors that can broadly be included in two main categories. In environmental sex determination (ESD), the process of establishment of the phenotypic sex can be modulated by external sources such as temperature at crucial stages of development (the thermosensitive period- TSP) in groups like reptiles (reviewed in Janzen & Philips, 2006) and much more rarely in fish (reviewed in Ospina-Alvarez & Piferrer, 2008), local sex ratio (Shapiro, 1980), or even population density (Munday *et al.*, 2006). Less plasticity is observed in genetic sex determination (GSD) in which mechanisms are mediated by strictly regulated gene cascades (Whilhem *et al.*, 2007). This type of sex determination is common to many species and involves the initial action of at least one central regulator gene (Zarkower, 2001), irrespective of the presence or absence of differentiated sex chromosomes (e.g. sex chromosomes have been identified in birds but no central regulator gene for sex determination has yet been proposed –for a review see Ellegren, 2001).

In fact, in genetic sex determination, triggering of one of two possible pathways can be accomplished by a diversity of processes that range from heterozigosity at a particular locus (Cook, 1993), the sex chromosome to autosome ratio (Penalva & Sanchez, 2003), to the presence of sex

chromosomes harboring a "master" regulator gene that initiates the cascade of events that culminates with the establishment of one or the other sex (Wilhelm et al, 2007).

Despite the variety of mechanisms, the components of sex determination cascades are apparently conserved throughout the vertebrate lineage (Smith *et al.*, 1999; Zarkower, 2001; Schartl, 2004), although their position in gene hierarchies and their interaction patterns can vary according to group. The sex determination cascade of mammals is one of the most well understood and has helped to elucidate patterns of gene interaction that are useful as a comparison basis for other groups (Wilhelm *et al.*, 2007). In mammals, the expression of the testis-determining factor on the Y chromosome, Sry, triggering male development, initiates a regulatory network of genes that modulate testis formation, which are initially antagonized by Dax1 expression in females (Fig. 2). However, the top of the cascade seems to be highly variable between groups, and the central role of Sry as the master male determining gene is only restricted to Eutherian mammals, making a strong point to the hypothesis of an overall conservation of downstream effectors of sex determination cascades as opposed to a considerable diversity of more upstream primary determinants (for a review see Zarkower, 2001).

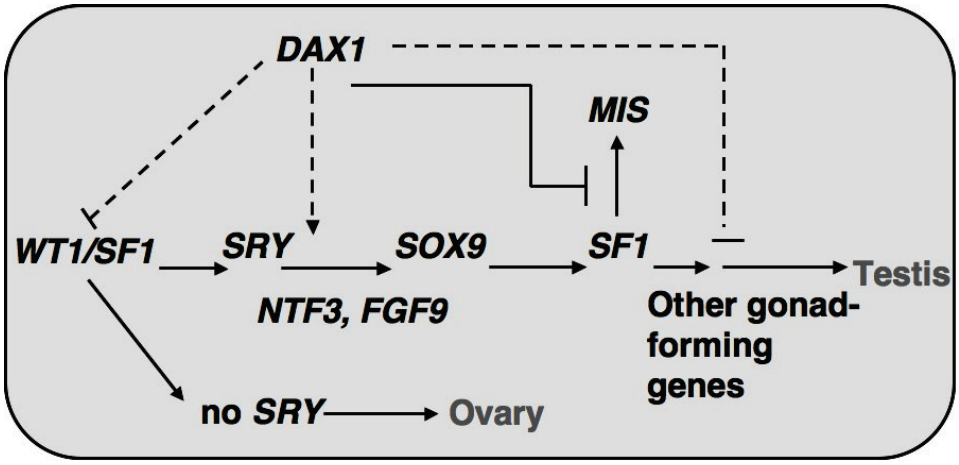


Figure 2. Initial cascade of molecular events taking part in mammalian sex determination. Lines in bold represent interaction for which experimental evidence has been gathered; dashed lines represent theoretically predicted pathways. Adapted from Ludbrook & Harley, 2004.

One paradigmatic example regarding the variability of the hierarchical positions of such genes across vertebrates is the *Dmrt1* gene. Although it is a downstream effector in mammalian male sex determination, *Dmrt1* plays a key role in sexual development of such diverse groups as birds, reptiles and fish **(**Ferguson-Smith, 2007). It has in fact been implicated as a good candidate for central male-determining gene in birds, due to its location in the Z chromosome and its apparent escape from

dosage compensation (Nanda *et al.*, 2000; Shan *et al.*, 2000). Comparative studies report an overall proximity of downstream effectors and a larger variability of the central genes in determination between different taxa. It has been proposed that genes with ancient roles in sexual regulation should have been retained in different taxa and that differential selective pressure in distinct groups has led to the addition of different upstream regulators (Marín & Baker, 1998).

1.3.2. Sex determination in fish

The variability of sex determination processes has been demonstrated among different vertebrate taxa, but no group offers so much diversity to the study of the evolution of sex determination as teleost fish. In this group, a complete set of sexuality systems can be observed, ranging from sequential hermaphroditism (in which an individual can develop as a male and subsequently as female -protandrous- or vice versa - protogynous) to gonochorism (in which the two different sexes are established in independent individuals). Among gonochoristic fish, sex determination can be modulated by a broad diversity of mechanisms, ranging from environmental to genetic sex determination (Devlin & Nagahama, 2002; Volff *et al.*, 2007).

Most of the knowledge of sex determination cascades in this group comes from model fish, but over the years further exploring of the complex evolution of sex development in fish has been achieved with contributions from different species (Devlin & Nagahama, 2002). Unlike in mammals, the major sex determination genes have not yet been described in most fish species, and it is quite likely that there might not be an uniform regulation pattern and that different genes will be proven to be the upstream regulator of sex determination in distinct species (Tanaka *et al.*, 2007). One of the better-characterized systems in terms of sex determination in fish is the medaka *Oryzias latipes*. It has male heterogametic sex determination system, in which, unlike mammals, sex chromosomes are morphologically indistinguishable (Matsuda *et al.*, 1998). Nevertheless, sex reversals can be induced by temperature alterations or hormone treatment. It is also the only non-mammalian species for which a master sex-determining gene has been identified. This gene (*dmrt1bY* or *dmy*) is a Y-specific copy of the autosomal *dmrt1* gene and its occurrence as the only functional gene in Y-specific region, the observation of sex reversal in case of mutations in *dmrt1bY* and the maintenance of its expression patterns after estrogen treatment were all consistent with an upstream role in the medaka sex differentiation (Nanda *et al.*, 2002; Matsuda *et al.*, 2002).

Conversely, in the cyprinid *Danio rerio*, one of the most important non-mouse models, there is a general lack of information regarding the mode of sex determination. No sex chromosomes were described in zebrafish and, despite the evidence of a strong genetic basis (von Hofsten & Olsson,

2005), no central regulator of sex determination has been described until now. The zebrafish *D. rerio* also presents an additional characteristic that might raise some difficulties in the study of sex determination: the occurrence of a type of gonadal development that is designated as juvenile hermaphroditism. In zebrafish larvae, an initial ovary-like structure is established, undergoes apoptosis in half of the fish and testes are subsequently developed from the undifferentiated gonad parts (Uchida *et al.*, 2002). The limited knowledge of sex determination and the process of sex differentiation extends to other model species such as the pufferfish *Tetraodon nigroviridis* and the fugu *Takifugu rubripes* (Li *et al.*, 2002), although a female heterogametic determination system, based on a sex-linked marker, has been recently proposed for the latter species (Cui *et al.*, 2006). In *Xiphophurus maculatus*, a fish model for cancer research, sex determination involves the action of three sex chromosomes: X, Y and W. Males can be XY or YY and females XX, XW or YW. Different models have been proposed to account for the sex determination in this species, including autosomal repression of X and W male-determining genes, exclusive activation of the Y alleles and the possibility of occurrence of dosage effects resulting from the presence of different copy numbers of the master male determination gene (2 copies in the Y, 1 copy in the X and no copy in the W chromosome). However, the complete genetic basis underlying these processes is yet to be uncovered (reviewed in Volff *et al.*, 2007).

In non-model species, the search for elements acting on the sex determination process and the attempt to elucidate the genetic network and regulatory mechanisms underlying sexual development has also been conducted in “famous” fish species like the nile tilapia *Oreochromis niloticus* (Guan *et al.*, 2000; Wang *et al.*, 2002) and the rainbow trout *Oncorhynchus mykiss*, (Marchand *et al.*, 2000; Brunelli *et al.*, 2001). For most species, however, a complete overview of the basis of sex determination has not been accomplished.

Evidence regarding the composition and hierarchical regulation of sex determination cascades is even scarcer when considering hybrid fish lineages. Difficulties arise because hybridisation events often result in a disruption of the sex determination patterns exhibited by parental species, usually culminating in a strong deviation of sex ratio distribution. The hybrid *P. formosa* could serve as an example to illustrate this phenomenon. Several poeciliid species essentially show a male heterogametic system, in which differential degeneration of male sex chromosomes is observed according to species (Traut & Winking, 2001). Conversely, a female heterogametic system compatible with the ZW/ZZ type has been reported for the *P. formosa*’s ancestor *Poecilia latipinna* (Sola *et al.*, 1990). The genetic framework underlying sex determination has not been elucidated, but the recent origin (5 to 30 MYA) and ongoing evolution of the poeciliid group of fish makes it very unlikely that one unique master regulator should be accounted for the process in all species (Volff &

Schartl, 2001). Conversely, as a result of hybridisation, all genes specifying the male phenotype were apparently "shut-down" in all-female *P. formosa*. The male development "program" is however still present in the hybrids, as phenotypic males can be induced by androgenic steroid treatment (Schartl *et al.*, 1991), implying an abrupt shift in the regulatory network upon hybridisation.

Hence, additional nuance regarding the question of how sex determination cascades are regulated in different contexts is brought upon by hybrid fish species. Globally, the mechanisms for the establishment of the phenotypic sex are widely diverse and probably represent distinct evolutionary pathways and in fish, more than in any other group, we observe that as far as sex determination is concerned, there is more than one means to an end.

1.4. Sex determination-II: players

Over the years, several evolutionary conserved genes have been analysed in teleost fish, in an attempt to identify specific roles and relative importance in the sexual development. The diversity of gene expression patterns and putative roles has impeached a global prediction of the interactions between players within the fish group. Genes such as *sox9* and the anti-Müllerian hormone *amh* (or *mis*) have been isolated and characterized in zebrafish *D. rerio* and in the medaka *O. latipes* (Rodríguez-Marí *et al.*, 2005; Klüver *et al.*, 2007) (Fig. 3). *Dmrt1*, a gene with downstream expression in mammals, has been implicated in key events of sex determination and gonad differentiation in the medaka (Kobayashi *et al.*, 2004), the nile tilapia *O. niloticus* (Guan *et al.*, 2000) and in the rainbow trout *O. mykiss* (Marchand *et al.*, 2000). *Wt1* has been isolated in the medaka (Fedorova *et al.*, 2008), zebrafish (Bollig *et al.*, 2006), the Japanese eel *Anguilla japonica* (Nakatsuru *et al.*, 2000) and the rainbow trout *O. mykiss*, (Brunelli *et al.*, 2001) and possible interactions of this gene with *sf1*

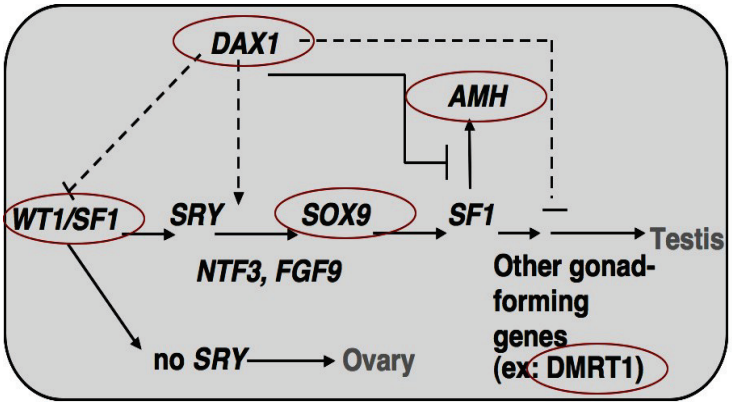


Figure 3. Examples of genes implicated in mammalian sex determination (adapted from Ludbrook & Harley, 2004), reported to have roles, at different hierarchical levels, in fish sex determination (highlighted in red)

have been reported in zebrafish (Hsu *et al.*, 2003). *Dax1* is reportedly involved in gonad development and differentiation in the nile tilapia *O. niloticus* (Wang *et al.*, 2002), the European sea bass *Dicentrarchus labrax* (Martins *et al.*, 2007) and the medaka (Nakamoto *et al.*, 2007).

The data gathered in these studies has further confirmed the idea that sex determination gene cascades, in an evolutionary point of view, present characteristics of robustness, by maintaining successful regulators over evolutionary time, but at the same time flexibility, as the same gene can occupy distinct hierarchical positions in the cascade, even in closely related groups. Sex determination in teleost fish is apparently no exception to this "rule" (Volff *et al.*, 2003).

1.5. The *Squalius alburnoides* complex

1.5.1 Origins

The *Squalius alburnoides* (Steindachner) fish complex was initially described by Collares-Pereira (1983, 1984), based on morphological and cytogenetic features. Two distinct forms were described, with similar haploid karyotype structure, but a differential number of gillrakers and pharyngeal teeth: form A, the most common (75-80%), composed mainly by triploids, with a pharyngeal tooth formula of 5-5 and 12-17 gillrakers, and form B, composed by diploids with a pharyngeal tooth formula of 4-4 (only rarely 5-5) and a larger number of gillrakers (17-26). A strong deviation of sex ratios towards females was observed in form A, and it was proposed to be composed of individuals with asexual reproduction (Collares-Pereira 1985, 1989).

Subsequent studies have shown that *S. alburnoides* has a widespread distribution not only in Portugal, but among the whole Iberian Peninsula and provided further insight into the events that gave rise to this complex of cyprinids. The first molecular evidence of a hybrid origin came from the report of fixed heterozygosity at numerous allozyme loci, analysed over several populations of *S. alburnoides*. A complete set of alleles was indistinguishable from the ones of sympatric *Squalius* species while the other could not be attributed to any other species within the distribution of the complex (Alves *et al.*, 1997a; Carmona *et al.*, 1997). A more precise overview of the ancestry of the hybrid complex was elucidated by the analysis of mitochondrial DNA sequences, in which the similarity between the *S. alburnoides* and the *Squalius pyrenaicus* (Günther) cytochrome b gene, implicated the later species as the maternal ancestor of the complex (Alves *et al.*, 1997b). Thus, the hybridisation events that gave rise to the complex would have most likely resulted from crosses involving females of *S. pyrenaicus* and males of an unknown ancestor. The search for the paternal ancestor of the complex was therefore launched and over the years several authors have provided clues towards its identification, consensually proposing by different approaches that it should be

one species closely related to the "non-Squalius" *Anaecypris hispanica* (Steindachner), which had subsequently disappeared from the complex distribution range (Alves *et al.*, 2001; Robalo *et al.*, 2006; Gromicho *et al.*, 2006) (Fig. 4).

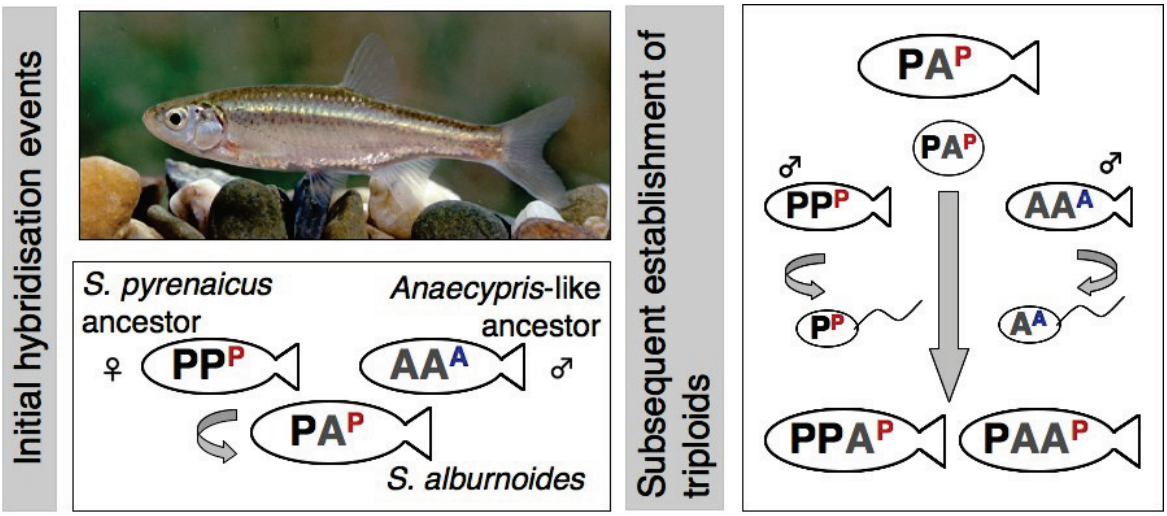


Figure 4. Representation of the hybridisation events that gave rise to the *S. alburnoides* complex. Possible initial trajectories for the establishment of the triploid forms are represented on the right side. Nuclear genomes are represented in black (P) and grey (A); mitochondrial genomes are represented in superscript letters in red (of *S. pyrenaicus*) and blue (of the paternal ancestor). Photo of *S. alburnoides* specimen by Isabel Catalão.

If the hybrid origin and the species putatively involved in the constitution of the *S. alburnoides* complex are no longer a matter of disagreement, a consensus has not been reached regarding the geographical location and the timing of these hybridisation events. Independent origins in the Tagus-Guadiana and the Sado River basins have initially been postulated (Alves *et al.*, 1997b). Based on the presence of *pyrenaicus*-type mitochondrial DNA in *S. alburnoides* outside the distribution range of this species, it was further proposed that the complex should have dispersed from the Tejo basin towards northern drainages by stream captures or inversions of flow direction (Alves *et al.*, 1997b). An alternative hypothesis of five independent origins of the complex has been proposed, based on the analysis of the complete cytochrome b gene, dating the hybridisation events in an extended time period, going back to the upper Pliocene (Cunha *et al.*, 2004). According to this hypothesis, three additional independent hybridisation sources would have established in Guadiana-Guadalquivir, Douro and in the Quarteira River. Therefore, in the northern region of the Iberian Peninsula (Mondego and Douro River basins), the complex should have been established by local hybridisation with *S. pyrenaicus*, with the latter species being subsequently wiped out from these locations by vicariant events (Cunha *et al.*, 2004), with the complex shifting to sperm dependency on the non-

parental taxon *S. carolitertii*. More recently, the hypothesis of an initial single origin followed by dispersion has been revived, based once more in the study of mitochondrial DNA variation. It has been postulated that the complex might have originated in the area of what is now the River Guadiana, less than 0.7 MY ago, and subsequently dispersed throughout the connections between river basins. According to this scenario, the Mondego and Douro Basins would have been colonized from Tejo, respectively at around 0.05 MY and 0.01 MY ago, through stream capture involving adjacent tributaries (Sousa-Santos *et al.*, 2007). Despite the obvious differences, the notion of differential establishment times, the distinctiveness in the evolutionary pathways and the establishment of different lineages within the complex are common to all proposed scenarios.

Presently, the complex has a widespread distribution in the Iberian Peninsula and occurs in sympatry with three bisexual species of the genus *Squalius*: the maternal ancestor species *S. pyrenaicus* in southern basins, *Squalius carolitertii* (Doadrio) in northern drainages and *Squalius aradensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira) in the small independent drainage of Quarteira (distribution in Portugal depicted in Fig. 5). The sympatry of *S. alburnoides* with *S. arandensis* has more recently been described and the complexity of the hybridisation processes in this location has not yet been fully unraveled. Additionally, there are reports of possible massive introgression from *S. aradensis* nuclear and mitochondrial genes into the complex in the Quarteira drainage (Sousa-Santos *et al.*, 2006a). Therefore, the works that compose this thesis focused mainly in the larger areas *of S. alburnoides* distribution, in which it occurs in sympatry with *S. pyrenaicus* and *S. carolitertii* and for which a more extensive knowledge of origin, population constitution (Fig. 5) and evolutionary features has already been obtained.

1.5.2. Forms and reproductive dynamics

One of the characteristic features of the *S. alburnoides* complex is its high diversity of reproductive modes that promote an intricate network of genetic exchange and a continuous shifting between forms[2]. The sympatric bisexual species *S. pyrenaicus* and *S. carolitertii* actively participate in the reproductive dynamics of *S. alburnoides*, by intercrossing with various forms of the complex and by contributing with new genetic material. As a result from this interaction, the P (of *S. pyrenaicus*) and C (of *S. carolitertii*) genomes are represented in hybrid individuals, according to geographical locations (P in southern and C in northern drainages).

[2] Each ploidy/genotype combination will be designated as “form” (e.g. in southern populations there are two triploid forms: PAA and PPA).

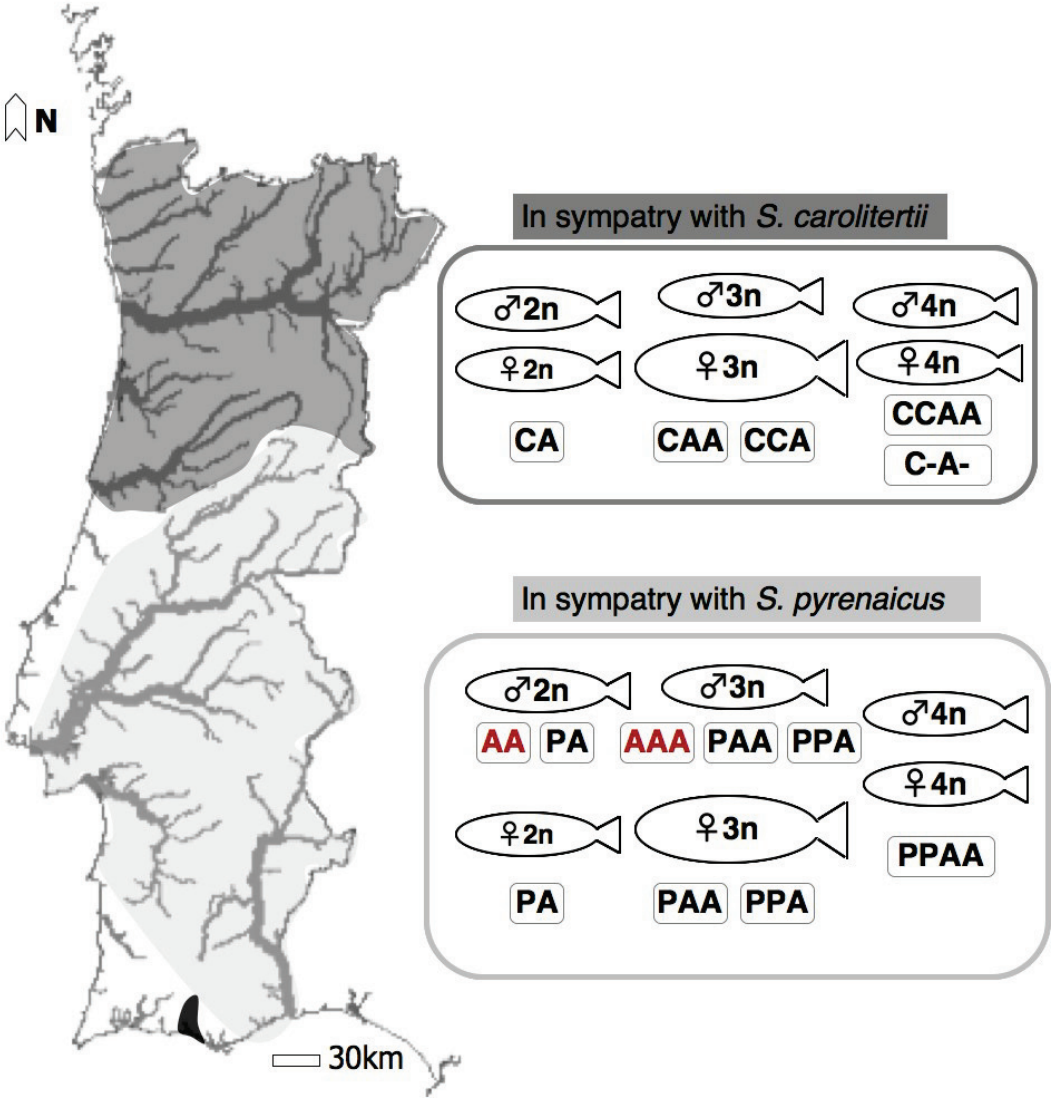


Figure 5. Distribution of the *S. alburnoides* in Portugal and the areas of sympatry with *S. carolitertii* (dark grey), *S. pyrenaicus* (light grey) and *S. aradensis* (black). The global composition of the populations of the complex in the two areas of sympatry with the bisexual *Squalius* species *S. carolitertii* and *S. pyrenaicus* are indicated, although local abundances can vary according to river. Forms that occur exclusively in southern populations are highlighted in red.

In the southern basins of Tejo and Guadiana the complex includes diploid (2n=50) (PA) and triploid (3n=75) (PAA, PPA) forms, with variable distribution of sex ratios (Fig. 3). Tetraploids (4n=100) are present in the Tejo Basin and apparently absent from Guadiana. In these locations the most predominant form is the triploid PAA female, although prevalence of different forms can vary locally (Alves *et al.*, 2001; Crespo-Lopéz *et al.*, 2007). Another locally abundant, morphologically distinct form has been reported exclusively in southern populations, corresponding to the form B of *S. alburnoides* initially identified by Collares-Pereira (1983, 1984). It presents AA or, more rarely, AAA genotypes, thus resembling the genome constitution of the paternal ancestor, but the presence of

pyrenaicus-like mitochondrial DNA implied that it has been reconstituted from the complex, thus being designated from then on as the "nuclear non- hybrid" form of *S. alburnoides* (reviewed in Alves *et al.*, 2001). This form has been reported as being constituted exclusively by males, with only one female individual reported (Sousa-Santos *et al.*, 2006b) among hundreds of individuals analysed since the 1980's.

Northern populations globally comprise the same type of forms, including diploids (CA), triploids (CAA and CCA) and tetraploids (CCAA, CAAA) (with the exception of "nuclear non-hybrid" specimens that have not been reported in northern locations), but some differences in terms of their occurrence and distribution were observed between the two main basins (Fig. 3). In the Mondego River basin, in which diploid CA individuals of both sexes were identified, triploid females (CAA) predominate although rare CCA individuals were also reported. There was no indication of occurrence of tetraploids in this basin (Pala & Coelho, 2005). In the Douro River Basin, the majority of the populations also show a clear triploid female predominance (CAA and CCA individuals were reported) and diploid and triploid individuals have also been identified (Carmona *et al.*, 1997). Recently, a distinctive distribution of forms has been reported in Rivers Lodeiro and Paiva within the Douro Basin: tetraploids (CCAA) were reported to be predominant, comprising 88% of all the individuals collected in the two locations. Triploid and diploid forms were also observed, but in a much lower proportion comparatively to other populations of the complex (Cunha *et al.*, 2008).

All the different forms of *S. alburnoides* interact through diverse reproductive modes that include strictly asexual mechanisms, but most commonly, pathways pending towards a more functional proximity to sexual reproduction, involving recombination, meiosis and gamete syngamy (Fig. 6). Diploid hybrid females and males of southern basins (PA) have been shown to be fertile and to transmit their genomes clonally to the progeny, producing diploid clonal eggs and sperm (Alves *et al.*, 1998, 1999). Only a very low proportion of diploid eggs develop by gynogenesis (<3%), and most commonly high levels of syngamy are observed. Clonal inheritance of paternal alleles was confirmed by crosses involving CA diploid males in the Mondego River basin (Pala & Coelho, 2005), implying similarities in the reproductive processes of diploid males throughout the geographical range of *S. alburnoides*. Conversely, Carmona *et al.* (1997) proposed that CA from the Douro River basin would reproduce by hybridogenesis, discarding the C genome during oogenesis (Fig. 7).

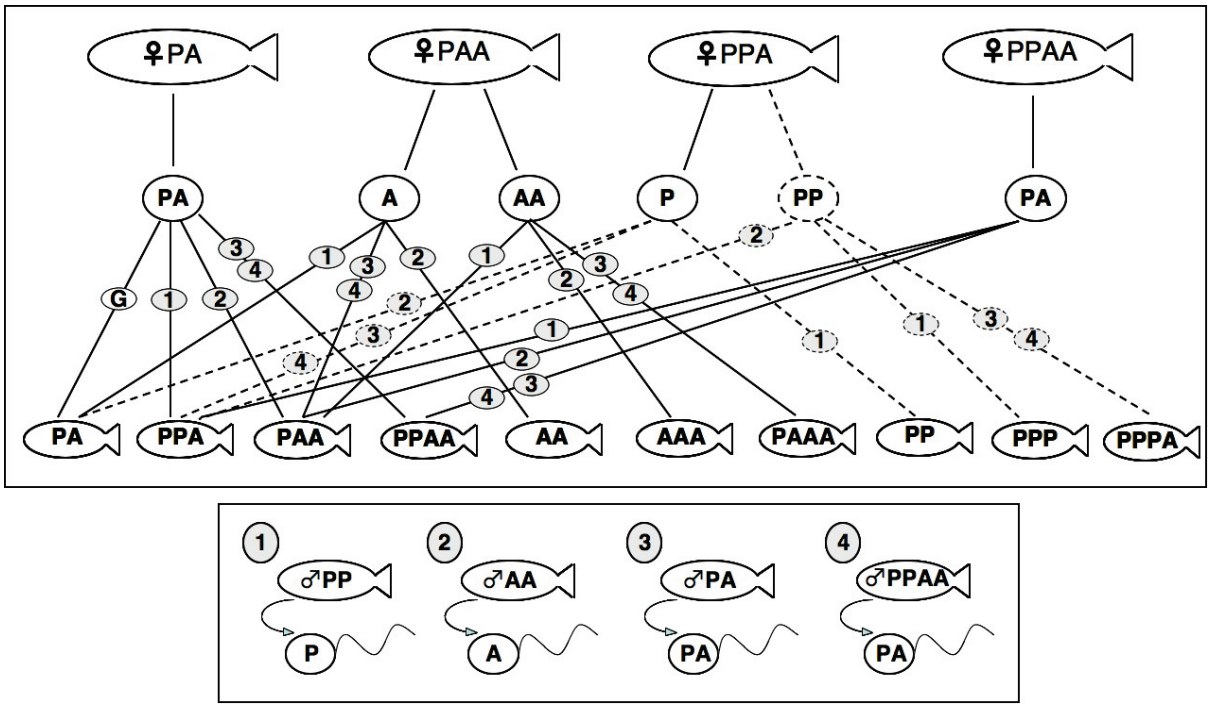


Figure 6. Summary of the putative reproductive modes of the different forms of the *S. alburnoides* complex in southern basins (adapted from Alves *et al.*, 2001; Crespo-Lopéz *et al.*, 2006). (—) Forms and pathways confirmed experimentally; (---) hypothetical pathways; (G) occurrence of gynogenesis; (1) fertilization by P sperm, produced by normal meiosis by *S. pyrenaicus*; (2) fertilization by A sperm, produced by normal meiosis by nuclear nonhybrid males; (3) fertilization by PA sperm produced clonally by diploid hybrid males; (4) fertilization by PA sperm produced by normal meiosis by tetraploid males.

Mechanisms of hybridogenesis (in which one parental genome is excluded from gametogenesis, while the other is transmitted to the egg without recombination) and, more commonly of meiotic hybridogenesis (in which the process of genome exclusion occurs like in hybridogenesis, but followed by recombination and meiosis between the remaining genomes) have been reported in the processes of gamete production of triploid PAA females (Alves *et al.*, 1998). A similar pathway of A genome exclusion and majority genome (P) meiosis has been proposed for PPA females, based on the result of an experimental cross (Crespo-López *et al.*, 2006). Moreover, the coincidence of two gamete formation processes in the same individual has been put into evidence in *S. alburnoides*, through the report of the simultaneous production of triploid and haploid eggs by a triploid (PAA) female (Alves *et al.*, 2004).

Triploid females of CAA genotype from the Douro basin have been shown to discard the C genome from oogenesis, through the study of allozyme markers, but the occurrence of a subsequent reductional process could not be assessed (Carmona *et al.*, 1997). More recently, the results of experimental crosses involving triploid CAA females from the Mondego River revealed that only one

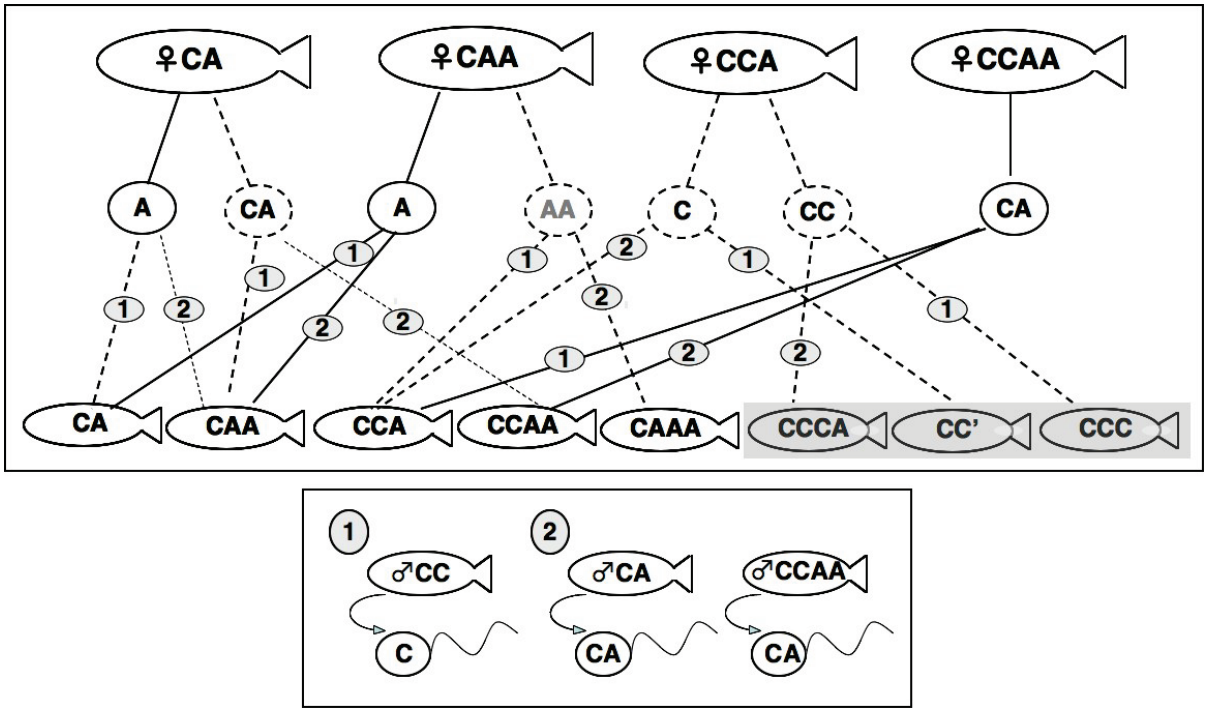


Figure 7. Summary of the putative reproductive modes of the different forms of the *S. alburnoides* complex in northern basins (adapted from Carmona *et al.*, 1997; Pala & Coelho, 2005; Cunha *et al.*, 2008) (—) Forms and pathways confirmed experimentally; (---) hypothetical pathways; (1) fertilization by C sperm, produced by normal meiosis by *S. carolitertii*; (3) fertilization by CA sperm either produced clonally by diploid hybrid males or though normal meiosis by tetraploid males. Forms that have not been observed in natural populations are represented in light grey.

set of A alleles was inherited by the progeny, implying that meiotic hybridogenesis processes might also be taking place in northern populations (Pala & Coelho, 2005). Regarding the remaining triploid forms, the production of unreduced sperm by triploid PAA and PPA hybrid males has been reported (Alves *et al.*, 1999), but the fertility and the mechanisms underlying gamete production of the CCA form (males and females) are yet to be assessed. Simetrical tetraploid males and females from southern and northern populations (PPAA and CCAA, respectively) have been shown to produce fertile diploid gametes, apparently through normal meiosis (Crespo-López *et al.*, 2006; Cunha *et al.*, 2008). Normal meiosis and the production of fertile sperm under Mendelian inheritance patterns have also been shown in the "nuclear non-hybrid" form of *S. albunoides* (Alves *et al.*, 1998, 1999).

As P and C alleles are usually brought into the complex from the two sympatric bisexual species, the reproductive modes that allow for the recombination of A alleles are highly relevant to the maintenance of the variability of *S. alburnoides* (Figs. 6 and 7). Thus, the meiotic hybridogenesis in triploid females and the gametogenesis patterns of "nuclear non-hybrid" which mimetize the

presence and genetic contribution of the bisexual "A" ancestor should be pointed out as important features for the perpetuation of the complex, for its evolutionary success and for the progression of the form shifting dynamics.

1.5.3. Evolutionary potential

The initial views of hybridisation and ploidy rise as unrefutable sentences of variation loss, lack of adaptative potential and probable extinction have long been overthrown by the evidence of successful ancient polyploid lineages (Dowling & Secor, 1997). Within this context, the *S. alburnoides* complex presents a number of features that place it as an example of viability and evolutionary success among polyploid hybrid taxa.

The southern populations of *S. alburnoides* are regarded as a paradigmatic denial of the idea of "dead-end" fate for hybrid polyploids. In Tejo and Guadiana River basins, the study of microsatellite variation has revealed high genetic diversity and confirmed the intricate pathways of genetic exchange that contribute to the continuous shifting and maintenance of diversity of forms (Crespo-López *et al.*, 2006, 2007). These distinctive reproductive mechanisms not only allowed the incorporation of the sexually produced P genome, but also the generation of new genetic material from within the hybrid forms, through "normal" meiosis and recombination, thus attaining an effective "escape" from a potentially "mandatory" asexual fate. In the Tejo River basin, yet another important feature in terms of evolutionary potential has apparently arisen: the occurrence of fertile symmetrical (PPAA) tetraploids of both sexes (Alves *et al.*, 1999; Crespo-López *et al.*, 2006), that would open the pathway for the return to normal meiosis and biparental reproduction, thus widening the prospects of evolution of this polyploid lineage. However, no proof of a self-sustaining population of tetraploids was obtained in this Basin (Crespo-López *et al.*, 2006).

On the other side of the spectrum we find the Mondego River populations. This *S. alburnoides* lineage showed a much lower genetic variability compared to southern populations and an overall scarcity of possibilities of enhancing it. Nuclear "non-hybrid" males are absent from these populations, thus restricting the opportunity for introducing new genetic variants of the A genome. Moreover, the pathways for attaining a new form of "bisexual" independence by tetraploidisation are also apparently impaired in the Mondego River basin (Pala & Coelho, 2005).

Conversely, in the Douro River basin yet another important evidence of the evolutionary prospects of the complex has recently been reported. In two rivers within the basin, the occurrence of variable and apparently well adapted symmetrical tetraploids of both sexes, comprising more than 80% of the population, opens a clear possibility of ongoing speciation occurring upon a hybrid non-sexual transition. The return to a balance genome context, the possibility of intercrossing between

individuals that produce gametes by normal meiosis and the perspective of occurrence of reproductive isolation, provide further strength to the idea of hybridisation and polyploidy as an important stepping stone for evolution and speciation. For the first time in *S. alburnoides* a strong direct support for this hypothesis has been obtained (Cunha *et al.*, 2008).

1.5.4. The hybrid genome jigsaw I - sex ratio, sex determination

One of the most commonly referred features of the *S. alburnoides* complex is the overall predominance of triploids (mostly PAA and CAA) in the majority of the populations and the strong correlation of this ploidy level with the female sex (male triploids are rare in all populations) (Alves *et al.*, 2001; Pala & Coelho, 2005; Crespo-López et al, 2007; Cunha *et al.*, 2008). The occurrence of triploids has been shown to be a common feature among several hybrid lineages (Dawley *et al.*, 1987; Janko *et al.*, 2007; Lampert *et al.*, 2008), but the correlation to the female sex does not seem to occur in every group (Spolsky & Uzzel, 1986; Felip *et al.*, 1999; Stöck *et al.*, 2002). *S. alburnoides* diploid and tetraploid forms, on the other hand, do not seem to exhibit such strong deviations and sex ratios show a better approximation to the normal bisexual distribution.

Regarding sex ratio deviation, a less common feature observed in the *S. alburnoides* complex, is the occurrence of a lineage composed exclusively by males of either AA or AAA genotype (reviewed in Alves *et al.*, 2001). The recent report of the occurrence of one diploid female of AA genotype (Sousa-Santos *et al.*, 2006b), does not seem to put the strong correlation at stake, as over the years the hundreds of AA and AAA individuals analysed at the morphological, molecular and functional level were all identified as males (Collares-Pereira 1983, 1984; Carmona *et al.*, 1997; Alves *et al.*, 1998, 1999, 2002; Crespo-López *et al.*, 2006, 2007).

The strong correlation between sex and genotype irrespective of geographical location and differential abundance of forms may indicate that rather than being modulated by environmental factors, the sex determination mechanisms *in S. alburnoides* might have a strong genetic basis. Considering this hypothesis, sex determination could the result of two possible mechanisms: a strictly based genetic action, in which several genes contribute to the cascade of events that leads to the determination of the sexual phenotypes (von Hofsten & Olsson, 2005) or the presence of distinct sex chromosomes, harboring the major sex determination gene (Graves, 2006).

Until now, no morphologically distinct chromosomes were found between sexes (Gromicho, 2006) in individuals of *S. alburnoides*. Chromosome heteromorphism that would suggest a ZWfemale/ ZZmale system has been reported for the bisexual sympatric species *S. carolitertii* and the maternal ancestor *S. pyrenaicus* (Collares-Pereira *et al.* 1998), but later studies failed to find such well differentiated chromosome in both *Squalius* species (Gromicho, 2006). Data from experimental crosses revealed

that a ZW chromosomal system could not fully explain sex determination in *S. alburnoides* and that in addition to female determinants on the W chromosome, a minimum of one non-W-linked gene would have to be expressed differently in hybrid and nonhybrid genome combinations to account for the results obtained. Moreover, the "strength" of these sex determining factors might be variable depending on populations and parental species (Alves *et al.* 1998). Thus, it is highly probable that even if functionally some linkage groups could be directly related to the process of sex determination, morphological differentiation of sex chromosomes has not been attained in the *Squalius* species.

The hypothesis of a genetic based mechanism of sex determination and the identification of how players interact in the regulatory process is therefore a strong possibility to be addressed in *S. alburnoides*. An additional interest of exploring this perspective is raised by the deviations in *S. alburnoides* sex ratios, apparently brought upon by hybridisation and heterogeneous genome combination. It has been proposed that the conservation of regulatory patterns in vertebrates could imply a general regulation scheme, in which particular elements of hormonal regulation, dosage compensation and epigenetic phenomena could be incorporated in particular groups (Lavranos *et al.*, 2006). It is possible that such regulatory processes would face significant disturbance, upon abrupt changes in genomic context, such as hybridisation and ploidy rise. The balance in the distribution of sex ratios attained in the bisexual ancestor *S. pyrenaicus*, has apparently been lost in the hybrids, making it a very interesting system to start studying these regulatory disturbance phenomena. However, the molecular basis underlying the disruptions on the process of sex determination and sex ratio distribution has never been addressed in *S. alburnoides*.

1.5.5. The hybrid genome jigsaw II - genome interactions

As previously referred in this chapter, one of the most prominent questions brought upon by hybridisation and ploidy rise events is how two genomes that were previously under independent regulatory "agendas" react to being gathered within the same nucleus (Riddle & Birchler, 2003). Often, the reaction to these genome "shock" events involves epigenetic phenomena, gene silencing and unorthodox patterns of gene expression (Comai, 2000; Shaked *et al.*, 2001; Adams *et al.*, 2003).

The functional impact of hybridisation has never been directly addressed in *S. alburnoides*. In terms of inheritance, the genomes merged in the hybrids apparently act as independent genetic units and there has been only one observation (Alves *et al.*, 1997a) that could eventually be justified by inter-genomic recombination, among other more probable causes.

The only evidence indicating that the two gene hierarchies brought together in the hybrid might be recognized and acted upon differentially is the process of exclusion of the heteromorphic genome in the meiotic hybridogenesis of triploid females of *S. alburnoides* (Alves *et al.*, 1998). However, the exact mechanism by which this preferential exclusion is consubstantiated is yet to be determined.

1.6. Aims and structure of the thesis

The main aim of this thesis was to provide an initial understanding, on a vertebrate context, of the effects of hybridisation and polyploidy at the gene expression level. The starting point was a very desirable hybrid fish system in which many evolutionary questions had already been addressed but in which virtually nothing was known about the impact of hybridisation on gene expression, genome functionality and on the regulation of gene interaction cascades: the *S. alburnoides* complex. This thesis apparently addresses the question at two very distinct levels, comprising restricted processes as well as global effects, but in fact it accounts for and aims at providing new shades for the discussion of a more general evolutionary question: how does a hybrid system, comprising multiple genome copies function and what are the impacts of gene "behavior" in the persistence and success of polyploid taxa?
Five specific aims were established as the basis for this dissertation:

1. To identify and isolate candidate genes for sex determination and genes with more widespread expression (housekeeping genes) in *S. pyrenaicus and S. alburnoides*. The aim would be to create a molecular working basis for the study of two different impact levels of hybridisation: the effects on a specific process (sex determination) and the more global impact in the hybrid organism as a whole.

2. To characterize the patterns of expression of genes correlated to sex determination in the adult and during development, in both the parental species and the hybrids.

3. To compare the expression patterns of bisexual *S. pyrenaicus* and several hybrid forms of *S. alburnoides* in order to detect differences that might account for the disparities in sex ratio distribution observed in the hybrids.

4. To characterize the expression patterns of housekeeping and gonad specific genes in terms of genome specific contribution, in hybrids of *S. alburnoides*.

5. Correlate the expression data with the evolution of a particular system (*S. alburnoides*) and integrate the findings in the more global context of success and evolutionary potential of polyploids.

This thesis is composed by four distinct chapters (Chapters 2 to 5), corresponding to a number of publications in which all these questions are addressed. The first objective is represented in all chapters, as it constitutes the underlying basis that enables all further analyses. Different sets of a total of eleven tissue specific and housekeeping genes were isolated from *S. pyrenaicus* and *S. alburnoides,* and different perspectives of their expression patterns were investigated in each publication. Chapter 2 and 3 focus on the basis of sex determination (objective 2), through the analysis of candidate genes regarding their participation in sex determination cascades in the Iberian complex. In the first one, corresponding to the manuscript "Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*" published in Gene, the first of the candidates is analysed, characterized and its expression correlated to gonad development. In Chapter 3 "Candidate genes for sex determination in the *Squalius alburnoides* complex: an initial characterization of sex cascade elements in the context of a hybrid genome", four new genes were isolated and their features as players in the sex determination process were inspected by various methods, including *in situ* hybridisation and RT-PCR analysis. The comparison between the adult and developmental expression patterns of all genes in *S. pyrenaicus* and the *S. alburnoides* hybrids (objective 3) was performed both in Chapter 2 and 3, and further extended to the roles described for these genes in different teleost species in an attempt to better characterize their functionality.
Chapters 4 and 5 tackled objectives 4 and 5 in distinct but intertwining perspectives. In Chapter 4 "Dosage compensation by gene copy silencing in a triploid hybrid fish", published in Current Biology, the report of gene regulation through genome-specific allele silencing is made for the first time in a polyploid vertebrate, based on qualitative and quantitative expression analysis of six genes in triploids from southern populations of *S. alburnoides*. In Chapter 5 "Gene expression regulation and lineage evolution: The North and South tale of the *Squalius alburnoides* complex", the analysis of the contribution of each genome to overall gene expression was extended to northern populations of the complex and an heterogeneity of patterns was observed in relation to geographical origin. The evolutionary implications of such differences were extensively discussed in light of the current knowledge of *S. alburnoides* and other polyploid taxa and several hypotheses correlating gene expression to evolutionary fate were postulated. Finally, the various evolutionary and functional implications of the global results obtained in this thesis were debated in the Discussion (Chapter 6), in which they were addressed in the broader context of polyploid and hybrid taxa. The last part of the present dissertation (Chapter 7) corresponds to the enunciation of the main achievements and the various new questions that this work has put forward, as well as some remarks regarding topics

that should be addressed in the future studies of sex determination and genome regulation dynamics in the *S. alburnoides* complex. Two additional appendices were included after Chapter 7, corresponding to technical notes about the isolation of additional genes that, despite not being analysed in the remaining chapters, were part of the global approach described in this thesis. The results obtained for these genes were relevant for the global discussion and briefly reported in Chapter 6, justifying their inclusion as supplementary information.

1.7. References

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U S A. 100, 4649-4654.

Adams, K.L., Percifield, R., Wendel, J.F., 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168, 2217-2226.

Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. Curr Opin Plant Biol. 8, 135-141.

Adams, K.L., 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. J Hered. 98, 136-141.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 1997a. The *Rutilus alburnoides* complex (Cyprinidae): evidence for a hybrid origin. J. Zoolog. Syst. Evol. Res. 35, 1-10.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., Dowling, T.E., 1997b. Maternal ancestry of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae) as determined by analysis of cytochrome b sequences. Evolution 51, 1584 – 1592.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 1998. Diversity in the reproductive modes of females of the Rutilus alburnoides complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. Mol Biol Evol. 15, 1233-1242.

Alves, M.J., Coelho, M.M., Próspero, M.I., Collares-Pereira, M.J., 1999. Production of fertile unreduced sperm by hybrid males of thetes Rutilus alburnoides complex (Teleostei, Cyprinidae): an alternative route to genome tetraploidization in unisexuals. Genetics 151, 277-283.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111, 375 – 385.

Alves, M.J., Collares-Pereira, M.J., Dowling, T.E., Coelho M.M., 2002. The genetics of maintenance of an all-male lineage in the Squalius alburnoides complex. J Fish Biol. 60, 649–662.

Alves, M. J., Gromicho, M., Collares-Pereira, M. J., Crespo- López, E., Coelho, M. M., 2004. Simultaneous production of triploid and haploid eggs by triploid *Squalius alburnoides* (Teleostei: Cyprinidae). J. Exp. Zool. 301A, 552–558.

Arai, K., Mukaino, M., 1998. Electrophoretic analysis of the diploid progenies from triploid x diploid crosses in the loach *Misgurnus anguillicaudatus* (Pisces: Cobitidae). J. Exp. Zool. 280, 368–374.

Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H., Birchler, J.A., 2005. Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics 169, 389-397.

Beukeboom, L.W., Vrijenhoek, R.C., 1998. Evolutionary genetics and ecology of sperm-dependent parthenogenesis. J Evol Biol 11, 755–782.

Birchler, J.A., Veitia, R.A., 2007. The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell. 19, 395-402.

Blanc, G., Wolfe, K.H., 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667–1678.

Bogart, J.P., 1989. A mechanism for interspecific gene exchange via all-female salamander hybrids. Pp. 170-179. In Evolution and Ecology of Unisexual Vertebrates. RM Dawley & JP Bogart (Eds). New York State Mus. Bull. 466. New York.

Bollig, F., Mehringer, R., Perner, B., Hartung, C., Schäfer, M., Schartl, M., Volff, J.N., Winkler, C., Englert, C., 2006. Identification and comparative expression analysis of a second wt1 gene in zebrafish. Dev Dyn. 235, 554-561.

Brunelli, J.P., Robison, B.D., Thorgaard, G.H., 2001. Ancient and recent duplications of the rainbow trout Wilms' tumor gene. Genome 44, 455-462.

Carmona, J.A., Sanjur, O.I., Doadrio, I., Machordom, A. Vrijenhoek, R.C., 1997. Hybridogenetic reproduction and maternal ancestry of polyploid Iberian fish: the *Tropidophoxinellus alburnoides* complex. Genetics 146, 983-993.

Chain, F.J., Ilieva, D., Evans, B.J., 2008. Duplicate gene evolution and expression in the wake of vertebrate allopolyploidisation. BMC Evol Biol. 8, 43.

Collares-Pereira, M.J., 1983. Estudo sistemático e citogenético dos pequenos ciprinídeos ibéricos pertencentes aos géneros Chondrostoma Agassiz, 1835, Rutilus Rafinesque, 1820 e Anaecypris. PhD. Thesis, University of Lisbon, Lisbon, Portugal.

Collares-Pereira, M.J., 1984. The "Rutilus alburnoides (Steindachner, 1866) complex" (Pisces, Cyprinidae). I. Biometrical analysis of some Portuguese populations. Arquivos do Museu Bocage (Sér. A) 2, 111-143.

Collares-Pereira MJ. 1985. The "Rutilus alburnoides (Steindachner, 1866) complex" (Pisces, Cyprinidae). II. First data on the karyology of a well-established diploid-triploid group. Arquivos do Museu Bocage (Sér. A) 3, 69-89.

Collares-Pereira MJ. 1989. Hybridisation in European cyprinids: evolutionary potential of unisexual populations. Pp. 281-288. In Evolution and ecology of unisexual vertebrates. RM Dawley & JP Bogart (Eds). New York State Mus. Bull.

Collares-Pereira, M.J., Próspero, M. I., Biléu, R.I., Rodrigues, E. M., 1998. *Leuciscus* (Pisces, Cyprinidae) karyotypes: transect of Portuguese populations. Gen Mol Biol. 21, 63-69.

Comai, L., 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol. 43, 387-399.

Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., Byers, B., 2000. Phenotypic instability and rapid gene silencing in newly formed Arabidopsis allotetraploids. Plant Cell 12, 1551-1568.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6, 836-846.

Cook, J., 1993. Sex determination in the hymenoptera. A review of models and evidence. Heredity 71, 421-435.

Crespo-López, M.E., Duarte, T., Dowling, T., Coelho, M.M., 2006. Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. Zoology (Jena) 109, 277-286.

Crespo-López, M.E., Pala, I., Duarte, T., Dowling, T.E., Coelho, M.M., 2007. Genetic structure of the diploid–polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. J. Fish Biol. 71, 423-436.

Cui, J.Z., Shen, X.Y., Gong, Q.L, Yang, G.P., Gu, Q.Q., 2006. Identification of sex markers by cDNA-AFLP in *Takifugu rubripes*. Aquaculture 257, 30-36.

Cunha, C., Coelho, M. M., Carmona, J. A., Doadrio, I. 2004 Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridisation events. Mol Ecol. 13, 2807–2817.

Cunha, C., Doadrio, I., Coelho, M.M., 2008. Speciation towards tetraploidization after intermediate processes of non-sexual reproduction. Philos Trans R Soc Lond B Biol Sci. 363, 2921-2929.

Dawley, R.M., Schultz, R.J., Goddard, K.A., 1987. Clonal reproduction and polyploidy in unisexual hybrids of Phoxinus eos and Phoxinus neogaeus (Pisces: Cyprinidae). Copeia 1987, 275-283.

Dawley, R.M., 1989. An introduction to unisexual vertebrates. In: Dawley, R.M., Bogart, J.P. (Eds.), Evolution and Ecology of Unisexual Vertebrates. New York State Museum, Albany, New York, USA, pp. 1 – 18.

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate, PLoS Biol. 3, e314.

Devlin, R.H., Nagahama, Y., 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208, 191-364.

Dowling, T.E., Secor, C.L., 1997. The role of hybridisation and introgression in the diversification of animals. Annu. Rev. Ecol. Syst. 28, 593-619.

Ellegren, H., 2001. Hens, cocks and avian sex determination. A quest for genes on Z or W? EMBO Rep. 2, 192-196.

Fedorova, S., Miyamoto, R., Harada, T., Isogai, S., Hashimoto, H., Ozato, K., Wakamatsu, Y., 2008. Renal glomerulogenesis in medaka fish, Oryzias latipes. Dev Dyn. 237, 2342-2352.

Felip, A., Zanuy, Z., Carrillo, M., Piferrer, F., 1999. Growth and gonadal development in triploid sea bass (Dicentrarchus labrax L.) during the first two years of age. Aquaculture 173, 387–397.

Ferguson-Smith, M., 2007. The evolution of sex chromosomes and sex determination in vertebrates and the key role of DMRT1. Sex Dev 1, 2-11.

Gallardo, M.H., González, C., Cebrian, I., 2006. Molecular cytogenetics and allotetraploidy in the red vizcacha rat, Tympanoctomys barrerae (Rodentia, Octodontidae). Genomics 88, 214–221.

Graf, J.-D., Polls Pelaz, M., 1989. Evolutionary genetics in the Rana esculenta complex. Pp. 289–301. In Evolution and Ecology of Unisexual Vertebrates. RM Dawley & JP Bogart (Eds). New York State Mus. Bull. 466, New York.

Graves, J.A., 2006. Sex chromosome specialization and degeneration in mammals. Cell 124, 901-914.

Gromicho, M., 2006. Genome polyploidisation in the Iberian fish complex Squalius alburnoides. A case study of evolution-in-action through hybridisation and non-sexuality: a cytogenetic approach. PhD. Thesis, University of Lisbon, Lisbon, Portugal.

Gromicho, M., Coelho, M.M., Alves, M.J., Collares-Pereira, M.J., 2006. Cytogenetic analysis of Anaecypris hispanica and its relationship with the paternal ancestor of the diploid-polyploid Squalius alburnoides complex. Genome 49, 1621-1627.

Guan, G., Kobayashi, T., Nagahama, Y., 2000. Sexually dimorphic expression of two types of DM (Doublesex/Mab-3)-domain genes in a teleost fish, the Tilapia (*Oreochromis niloticus*).Biochem Biophys Res Commun. 272, 662-666.

Günther, R., Uzzell, T., Berger, L. 1979. Inheritance patterns in *triploid Rana 'esculenta'* (Amphibia, Salientia). Mitt. Zool.Mus. Berlin 55, 35–57.

Guo, M., Davis, D., Birchler, J.A., 1996. Dosage effects on gene expression in a maize ploidy series. Genetics 142, 1349-1355.

Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., Robertson, D.L., 2007. All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol. 8, R209.

Hsu, H.J., Lin, G., Chung, B.C., 2003. Parallel early development of zebrafish interrenal glands and pronephros: differential control by wt1 and ff1b. Development 130, 2107-2116.

Hubbs, C. L. & Hubbs, L. C., 1932. Apparent parthenogenesis in nature, in a form ofish of hybrid origin. Science 76, 628 –630.

Janko, K., Bohlen, J., Lamatsch, D., Flajshans, M., Epplen, J.T., Ráb, P., Kotlík, P., Slechtová, V., 2007. The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis: Teleostei), and their ability to establish successful clonal lineages--on the evolution of polyploidy in asexual vertebrates. Genetica. 131,185-194.

Janzen, F.J., Phillips, P.C., 2006. Exploring the evolution of environmental sex determination, especially in reptiles. J Evol Biol. 19, 1775-1784.

Kashkush, K., Feldman, M., Levy, A.A., 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160, 1651-1659.

Kim, I.-S., Lee, J.-H., 2000. Hybridisation experiment of diploid triploid cobitid fishes, *Cobitis sinensis-longicorpus* complex (Pisces, Cobitidae). Folia Zool. 49, 17–22.

Klüver, N., Pfennig, F., Pala, I., Storch, K., Schlieder, M., Froschauer, A., Gutzeit, H.O., Schartl, M., 2007. Differential expression of anti-Müllerian hormone (amh) and anti-Müllerian hormone receptor type II (amhrII) in the teleost medaka. Dev Dyn. 236, 271-281.

Kobayashi, T., Matsuda, M., Kajiura-Kobayashi, H., Suzuki, A., Saito, N., Nakamoto, M., Shibata, N., Nagahama, Y., 2004. Two DM domain genes, *DMY* and *DMRT1*, involved in testicular differentiation and development in the medaka, *Oryzias latipes*. Dev Dyn. 231, 518–526.

Lamatsch, D. K., 2001 Molekulargenetische und zytologische Untersuchungen zur paternalen Introgression beim gynogenetischen Amazonenkärpfling, *Poecilia formosa*. PhD thesis, University of Wuerzburg.

Lamatsch, D. K., Nanda, I., Schlupp, I., Epplen, J. T., Schmid, M., Schartl, M., 2004. Distribution and stability of supernumerary microchromosomes on natural populations of the Amazon molly, *Poecilia formosa*. Cytogenet. Genome Res. 106, 189 –194.

Lampert, K.P., Schartl, M., 2008. The origin and evolution of a unisexual hybrid: Poecilia formosa. Philos Trans R Soc Lond B Biol Sci. 363, 2901-2909.

Lavranos, G., Angelopoulou, R., Manolakou, P., Balla, M., 2006. Hormonal and meta-hormonal determinants of sexual dimorphism. Coll Antropol. 30, 659-663.

Le Comber, S.C., Smith, C., 2004. Polyploidy in fishes: patterns and processes. Biol. J. Linnean Soc. 82, 431-442.

Lee, H.S., Chen, Z.J., 2001. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. Proc Natl Acad Sci U S A. 98, 6753-6758.

Leggatt, R.A., Iwama, G.K., 2003. Occurrence of polyploidy in fishes. Rev Fish Biol Fisher. 13, 237-246.

Li, Y., Hill, J. A., Yue, G. H., Chen, F., Orban, L., 2002. Extensive search does not identify genomic sex markers in Tetraodon nigroviridis. J Fish Biol. 61, 1314–1317.

Liu, B., Vega, J.M., Feldman, M., 1998. Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. Genome 41, 535-542.

Liu, B., Wendel, J.F.,2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43, 874-880.

Ludbrook, L., Harley, V.R., 2004. Sex determination: a “window” of DAX1 action. Trends Endocrinol Metab. 15, 116-121.

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151-1155.

Ma, X.F., Gustafson, J.P., 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidisation. Cytogenet Genome Res. 109, 236-249.

Mallet, J., 2007. Hybrid speciation. Nature 446, 279-283.

Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J., Bernot, A., Laudet, V., Guiguen, Y., 2000. DMRT1 expression during gonadal differentiation and spermatogenesis in therainbow trout, *Oncorhynchus mykiss*. Biochim Biophys Acta 1493, 180-187.

Marín, I., Baker, B.S., 1998. The evolutionary dynamics of sex determination. Science 281, 1990-1994.

Martins, R.S., Deloffre, L.A., Mylonas, C.C., Power, D.M., Canário, A.V., 2007. Developmental expression of DAX1 in the European sea bass, *Dicentrarchus labrax*: lack of evidence for sexual dimorphism during sex differentiation. Reprod Biol Endocrinol. 5, 19.

Matsuda, M., Matsuda, C., Hamaguchi, S., Sakaizumi, M., 1998. Identification of the sex chromosomes of the medaka, Oryzias latipes, by fluorescence in situ hybridisation. Cytogenet Cell Genet. 82, 257-262.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., Sakaizumi, M., 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417, 559– 563.

Munday, P.L., Wilson White, J., Warner, R.R., 2006. A social basis for the development of primary males in a sex-changing fish. Proc Biol Sci. 273, 2845-2851.

Nadeau, H., Sankoff, D., 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution, Genetics 147, 1259–1266.

Nakamoto, M., Wang, D.S., Suzuki, A., Matsuda, M., Nagahama, Y., Shibata, N., 2007. Dax1 suppresses P450arom expression in medaka ovarian follicles. Mol Reprod Dev. 74, 1239-1246.

Nakatsuru, Y., Minami, K., Yoshikawa, A., Zhu, J., Oda, H., Masahito, P., Okamoto, N., Nakamura, Y., Ishikawa, T., 2000. Eel WT1 sequence and expression in spontaneous nephroblastomas in Japanese eel. Gene 245, 245-251.

Nanda, I., Zend-Ajusch, E., Shan, Z., Grutzner, F., Schartl, M., Burt, D.W., Koehler, M., Fowler, V.M., Goodwin, G., Schneider, W.J., Mizuno, S., Dechant, G., Haaf, T., Schmid, M., 2000. Conserved Synteny Between the Chicken Z Sex Chromosome and Human Chromosome 9 Includes the Male Regulatory Gene Dmrt1: a Comparative (Re)View on Avian Sex Determination. Cytogenetics Cell Genetics. 89, 67-78.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., Schmid, M., Schartl, M., 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. Proc Natl Acad Sci U S A. 99, 11778 – 11783.

Nef, S., Parada, L.F., 2000. Hormones in male sexual development. Genes Dev. 14, 3075-3086.

Ohno, S., 1970. Evolution by Gene Duplication. George Allen and Unwin, London.

Orr, A. H., 1990. "Why polyploidy is rarer in animals than in plants" revisited, Am. Nat. 136, 759–770.

Ospina-Alvarez, N,, Piferrer, F., 2008. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. PLoS ONE 3, e2837.

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu Rev Genet. 34, 401-437.

Otto, S.P., 2007. The evolutionary consequences of polyploidy. Cell 131, 452-462.

Pala, I., Coelho, M. M., 2005. Contrasting views over a hybrid complex: between speciation and evolutionary "dead-end". Gene 347, 283–294.

Panopoulou, G., Poustka, A.J., 2005. Timing and mechanism of ancient vertebrate genome duplications - the adventure of a hypothesis. Trends Genet. 21, 559-567.

Parker, E.D. Jr, Walker, J.M., Paulissen, M.A., 1989. Clonal diversity in Cnemidophorus: ecological and morphological consequences. Pp. 72–86. In Evolution and Ecology of Unisexual Vertebrates. RM Dawley & JP Bogart (Eds). New York State Mus. Bull. 466, New York.

Penalva, L.O.F, Sanchez, L., 2003. RNA binding protein sex-lethal (sxl) and control of *Drosophila* sex determination and dosage compensation. Microbiol Mol Biol Rev. 67, 343-359.

Plötner, J., Klinkhardt, M., 1992. Investigations on the genetic structure and the morphometry of a pure hybrid population of Rana kl. esculenta (Anura, Ranidae) in North Germany. Zool. Anz. 229, 163–184.

Ramsey, J., Schemske, D.W., 2002. Neopolyploidy in flowering plants. Annu Rev Ecol Syst. 33, 589-639.

Riddle, N.C., Birchler, J.A., 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Genet. 19, 597-600.

Robalo, J.I., Sousa Santos, C., Levy, A., Almada, V.C., 2006. Molecular insights on the taxonomic position of the paternal ancestor of the Squalius alburnoides hybridogenetic complex. Mol Phylogenet Evol. 39, 276-281.

Rodríguez-Marí, A., Yan, Y.L., Bremiller, R.A., Wilson, C., Cañestro, C., Postlethwait, J.H., 2005. Characterization and expression pattern of zebrafish Anti-Müllerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. Gene Expr Patterns. 5, 655-667.

Schartl, M., Schlupp, I., Schartl, A., Meyer, M.K., Nanda, I., Schmid, M., Epplen, J.T., Parzefall, J., 1991. On the stability of dispensable constituents of the eukaryotic genome: stability of coding sequences versus truly hypervariable sequences in a clonal vertebrate, the amazon molly, Poecilia formosa. Proc Natl Acad Sci U S A. 88, 8759-8763.

Schartl, M., Nanda, I., Schlupp, I., Wilde, B., Epplen, J.T., Schmid, M., Parzefall, J., 1995. Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish. Nature 373, 68–71.

Schartl, M., 2004. A comparative view on sex determination in medaka. Mech Dev. 121, 639–645.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., Levy, A.A., 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridisation and allopolyploidy in wheat. Plant Cell 13, 1749-1759.

Shan, Z., Nanda, I., Wang, Y., Schmid, M., Vortkamp, A., Haaf, T., 2000. Sex-specific expression of an evolutionarily conserved male regulatory gene, DMRT1, in birds, Cytogenet. Cell Genet. 89, 252–257.

Shapiro, D.Y., 1980. Serial Female Sex Changes After Simultaneous Removal of Males from Social Groups of a Coral Reef Fish. Science 209, 136-1137.

Simillion, C., Vandepoele, K., Van Montagu, M.C.E., Zabeau, M., Van de Peer, Y., 2002. The hidden duplication past of Arabidopsis thaliana, Proc. Natl. Acad. Sci. USA. 99, 13627–13632.

Sites, J.W. Jr, Peccinini-Seale, D., Moritz, C., Wright, J.W., Brown, W.M., 1990. The evolutionary history of the parthenogenetic Cnemidophorus lemniscatus (Sauria, Teiidae). I. Evidence for a hybrid origin. Evolution 44, 906–921.

Smith, C.A., McClive, P.J., Western, P.S., Reed, K.J., Sinclair, A.H., 1999. Conservation of a sex-determining gene. Nature 402, 601–602.

Sola, L., Monaco, P.J., Rasch, E.M., 1990.Cytogenetics of bisexual/unisexual *Poecilia* I. C-bands, AgNOR polymorphisms and sex chromosomes in three populations of *Poeciilia latipinna*. Cytogenet Cell Genet 53, 148-154.

Song, K., Lu, P., Tang, K., Osborn, T.C., 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci U S A 92, 7719-7723.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V.C., 2006a. Evidence of extensive mitochondrial introgression with nearly complete substitution of the typical *Squalius pyrenaicus*-like mtDNA of the *Squalius alburnoides* complex (Cyprinidae) in an independent Iberian drainage. Journal of Fish Biology 68 (Supplement B), 292–301.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V.C., 2006b. May a hybridogenetic complex regenerate the nuclear genome of both sexes of a missing ancestor? First evidence on the occurrence of a nuclear non-hybrid Squalius alburnoides (Cyprinidae) female based on DNA sequencing. Journal of Natural History 40, 1443–1448.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V., 2007. Reading the history of a hybrid fish complex from its molecular record. Mol Phylogenet Evol. 45, 981-996.

Spolsky, C., Uzzell, T., 1986. Evolutionary history of the hybridogenetic hybrid frog Rana esculenta as deduced from mtDNA analyses. Mol Biol Evol. 3, 44-56.

Stöck, M., Lamatsch, D.K., Steinlein, C., Epplen, J.T., Grosse, W.R., Hock, R., Klapperstück, T., Lampert, K.P., Scheer, U., Schmid, M., Schartl, M., 2002. A bisexually reproducing all-triploid vertebrate. Nat Genet. 30, 325-328.

Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., Sakaizumi, M., 2007. Evidence for different origins of sex chromosomes in closely related Oryzias fishes: substitution of the master sex-determining gene. Genetics 177, 2075-2081.

Traut, W., Winking, H., 2001. Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. Chromosome Res 9, 659-672.

Uchida, D., Yamashita, M., Kitano, T., Iguchi, T., 2002. Oocyte apoptosis during the transition from ovary-like tissue to testes during sex differentiation of juvenile zebrafish. J Exp Biol. 205, 711–718.

Volff, J.-N., Schartl, M., 2001. Variability of genetic sex determination in poeciliid fishes. *Genetica* 111, 101–110.

Volff, J.-N., Kondo, M., Schartl, M., 2003. Medaka *dmY/dmrt1Y* is not the universal primary sex-determining gene in fish. Trends Genet 19, 196-199.

Volff, J.-N., 2005. Genome evolution and biodiversity in teleost fish. Heredity 94, 280-294.

Volff, J.-N., Nanda, I., Schmid, M., Schartl, M., 2007. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev 1, 85-99.

von Hofsten, J., Olsson, P.E., 2005. Zebrafish sex determination and differentiation: involvement of FTZ-F1 genes. Reprod Biol Endocrinol. 3, 63.

Vrijenhoek, R.C., 1979. Factors affecting clonal diversity and coexistence. Amer. Zool. 19, 787-797.

Vrijenhoek, R.C., 1984. Ecological differentiation among clones: the frozen niche variation model. Pp. 217–231. In Population Biology and Evolution. K Wöhrmann & V Loeschcke (Eds). Springer-Verlag Berlin.

Wang, D.S., Kobayashi, T., Senthilkumaran, B., Sakai, F., Sudhakumari, C.C., Suzuki, T., Yoshikuni, M., Matsuda, M., Morohashi, K., Nagahama, Y., 2002. Molecular cloning of DAX1 and SHP cDNAs and their expression patterns in the Nile tilapia, *Oreochromis niloticus*. Biochem Biophys Res Commun. 297, 632-640.

Wendel, J.F., Schnabel, A. and Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). Proc. Natl. Acad. Sci. USA. 92, 280–284.

Wilhelm, D., Palmer, S., Koopman, P., 2007. Sex determination and gonadal development in mammals. Physiol Rev. 87, 1-28.

Zarkower, D., 2001. Establishing sexual dimorphism: conservation amidst diversity? Nat Rev Genet. 2, 75-85.

Zhang, Q.K., Arai, K., Yamashita, M., 1998. Cytogenetic mechanisms for triploid and haploid egg formation in the triploid loach *Misgurnus anguillicaudatus.* J. Exp. Zool. 281, 608–619.

CHAPTER 2

Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*

Irene Pala, Nils Klüver, Sólveig Thorsteinsdóttir, Manfred Schartl, Maria Manuela Coelho

Gene 410, 249-258 (2008)

Available online at www.sciencedirect.com
ScienceDirect
Gene 410 (2008) 249–258
GENE
www.elsevier.com/locate/gene

Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*

I. Pala [a,*], N. Klüver [b], S. Thorsteinsdóttir [a,c], M. Schartl [b], M.M. Coelho [a]

[a] *Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências de Lisboa, Lisbon, Portugal*
[b] *Physiologische Chemie I, Biozentrum, University of Würzburg, Würzburg, Germany*
[c] *Instituto Gulbenkian de Ciência, Oeiras, Portugal*

Received 10 September 2007; received in revised form 3 December 2007; accepted 12 December 2007
Available online 2 January 2008

Abstract

In fish of the *Squalius alburnoides* complex, hybridisation and polyploidy have affected sex ratios, resulting in strong correlations between sex and genotype. The preponderance of females among triploids and the occurrence of an all male lineage among diploids seem to imply that sex ratio deviations should have a strong genetic basis. Until now, no information has been gathered regarding the molecular basis of sex determination in this intricate hybrid system. Thus, putative regulatory elements of the cascade that potentially are involved in sex determination in *S. alburnoides* have to be investigated. Being reported to have an important role in teleost sex determination, and more particularly in male gonad development, the anti-Müllerian hormone, *amh* was a good initial candidate. Here we report the isolation, cloning and characterization of the amh ortholog in *S. alburnoides* and the ancestral species *S. pyrenaicus*. In adult S. alburnoides and *S. pyrenaicus* of both sexes, *amh* shows a gonad specific expression pattern, restricted to the Sertoli cell lineage in testis and to granulosa cells in ovaries. During development, it plays an early role in male gonad differentiation in *S. alburnoides*. Overall the observed patterns are similar to what has been reported in other teleost species. This suggests a conserved role of *amh* and implies that its expression dynamics cannot be directly responsible for the sex ratio deviations reported in *S. alburnoides*. It is possible that a conjunction of other factors could be contributing for sex ratio imbalance. The present results constitute the starting point in the characterization of the *S. alburnoides* sex determination cascade, a process that we expect to shed some light on the molecular basis of sex distribution, within the context of hybrid system evolution.

Keywords: cyprinids; hybridisation; sex ratio distribution; genetic sex determination; male gonad development

1. Introduction

Sex determination cascades have been extensively studied in higher vertebrates (Wilhelm et al., 2007) and the relative role of particular gene components has been widely reported. In other groups, such as teleost fishes, sex determination involves a broader set of mechanisms and is not always strictly dependent on genetic regulation (Barroiller et al., 1999). In teleosts, the position of distinct members of the sex determination cascade and their expression dynamics has only started to be investigated. So far only one sex determination master gene (*dmrt1bY* or *dmy*,) has been identified in a teleost species (Nanda et al., 2002; Matsuda et al., 2002). One of the genes that has recently been characterized as an important contributor to the regulation of sex determination in several teleost species like zebrafish (Rodriguez-Mari et al., 2005), medaka (Klüver et al., 2007), japanese eel (Miura et al., 2002), japanese flounder (Yoshinaga et al., 2004) and sea bass (Halm et al., 2007) is the *anti-Müllerian hormone* (*amh*). *Amh* is a member of the Transforming Growth Factor β (TGF-β) family and an important regulator in both early sex determination and later gonad development in higher vertebrate species (Josso and Clemente, 2003). In

Abbreviations: amh, *anti-Müllerian hormone*; cDNA, DNA complementary to RNA; dah, days after hatching; RACE, Rapid Amplification of cDNA Ends; TGF-β, Transforming Growth Factor β.

* Corresponding author. Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Bloco C2-3° piso, Campo Grande, 1749-016 Lisboa, Portugal. Tel.: +351 21 7500000 22312; fax: +351 217500028.
E-mail address: iapala@fc.ul.pt (I. Pala).

0378-1119/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.12.018

mammals it plays a crucial role in male sex determination, and is one of the first genes to be strongly expressed in the Sertoli cell lineage, being responsible for Müllerian duct regression during testicular differentiation (Josso et al., 1998). In females, its expression is weaker, established later and usually confined to the granulosa cell lineage (Durlinger et al., 2002).

Sex specific *amh* expression has been reported during gonad development in several species of teleosts (Miura et al., 2002; Rodriguez-Mari et al., 2005; Yoshinaga et al., 2004). The described expression pattern is consistent with a specific role for *amh* in the development of the Sertoli cell lineage and testicular differentiation, although expression is also detected in adult gonads of both sexes (Rodriguez-Mari et al., 2005). In other species, like the medaka (Klüver et al., 2007) no sexually dimorphic *amh* expression has been described, and *amh* seems to contribute to both male and female gonad development. Despite differences, *amh* expression patterns in most species studied so far share common features in terms of sex-and cell type-specific timing of expression. Thus it is an important gene to study with respect to understanding the sex determination process, in particular the establishment of the male phenotype, in teleosts (Wang and Orban, 2007). In addition, investigating the role of *amh* in the determination of a specific gonad type can be extremely useful for elucidating the mechanisms behind alterations in the normal distribution of sex ratios observed in some fish species.

The *Squalius alburnoides* fish complex is a very interesting example of altered sex ratios among different genotypes. This complex of cyprinid fish of hybrid origin is composed of fertile forms of various ploidy levels. The complex exhibits a set of unusual characteristics regarding the sex ratio distribution according to genotype and disruption of normal bisexual reproductive modes (Alves et al., 2001; Crespo-López et al., 2006). In *S. alburnoides*, male and female gametogenesis have been altered relative to their bisexual ancestors, *Squalius pyrenaicus* (maternal, P genome) and an *Anecypris hispanica*-like species (paternal, A genome), respectively. Diploid hybrids of *S. alburnoides* (PA) of both sexes produce unreduced gametes that in most cases give rise to non-clonal progeny through syngamy. In the majority of cases, triploid hybrid females produce haploid oocytes by "meiotic hybridogenesis" in which the genome from *S. pyrenaicus* is excluded during oogenesis. Random segregation and recombination occurs between the two non-excluded genomes. More rarely, diploid eggs are produced (reviewed in Alves et al., 2001). The complex exhibits a biased sex ratio, with a clear preponderance of females among triploids (particularly within the PAA form), and local increased abundance of males among (PA) diploids (Alves et al., 2001). A striking peculiarity of the complex is the occurrence of a lineage composed exclusively of males, corresponding to the nuclear genotype (AA), producing fertile sperm through normal meiosis (Alves et al., 2002). The question arises whether sex determination mechanisms have also endured disruptions that could explain the observed deviations in sex ratios. The relationship between genotypic combination and sex could suggest that genes involved in sex determination would – at least to some extent–be affected by hybridisation and genome interactions. In fact, disruption of normal expression patterns of some genes involved in sex determination has been reported in *Drosophila* species hybrids (Michalak and Noor, 2004). In the case of *S. alburnoides* hybrids, there is a complete lack of knowledge regarding sex determination and no gene potentially participating in the process has ever been isolated. In addition, no information is available about the mechanism of sex determination in the maternal bisexual ancestor of the complex, *S. pyrenaicus*. Thus, at present, it is impossible to study sex determination in the *S. alburnoides* complex using a comparative approach.

Taking into account that *amh* is one of the genes that has been reported to actively participate in the pathway of male sex determination in several teleost species, and considering that in *S. alburnoides* sex ratios are biased by unknown reasons, it is important to determine markers for the male and female lineages in this complex. In this study, we have isolated the *amh* gene and characterised its expression patterns in *S. alburnoides* and the ancestral bisexual species *S. pyrenaicus*. With this approach we aimed to provide a first insight into the characterization of one of the putative elements of the sex determination cascade of this intricate system. We show that *amh* is expressed in male embryos and adults in a similar way as in other teleost species. *Amh* expression is a good marker for male embryos of *S. alburnoides* and can be used in future studies addressing the sex determination mechanism of the complex.

2. Materials and methods

2.1. Collection of samples

A total of 20 specimens, 6 of *S. pyrenaicus* and 14 of *S. alburnoides* (7 nuclear non-hybrid, 2 diploid and 5 triploid hybrids) were collected from the Tejo River Basin (River Raia), where all forms that compose the complex are putatively present. All individuals were brought alive to the laboratory, identified morphologically and selected according to gonad maturity (verified through light abdominal pressure) for experimental crosses. A total of 13 individuals of both *S. pyrenaicus* (4) *and S. alburnoides* (9) were sacrificed in order to obtain adult female and male gonads for RNA extraction. Fin clips were collected from each individual, stored in ethanol at 4 °C and were later used for DNA extraction. All other samples were stored at −80 °C.

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted from adult gonads of *S. pyrenaicus* and *S. alburnoides* using the TRIZOL reagent (Gibco-BRL) according to the supplier's recommendation. First strand cDNA was synthesized with RevertAid™ First Strand cDNA Synthesis Kit (Fermentas), using random hexamer primer.

2.3. Ploidy determination

Blood samples were drawn from the caudal vein, stabilized in buffer (40 mM citric acid trisodium salt, 0.25 M sucrose, and 5% dimethyl sulfoxide) and immediately frozen at −80 °C.

Flow cytometry measurements were conducted as described in Dawley & Goddard (1988).

2.4. DNA extraction and genotyping of specimens

From each adult *S. pyrenaicus* or *S. alburnoides* specimen from which gonads were collected, total genomic DNA was obtained following standard digestion protocols with SDS and proteinase K. DNA extraction was performed using an alternative method to the phenol/chloroform protocol adapted from Miller et al. (1988). Genotypes were determined by three informative microsatellite loci previously used in the study of Southern populations of the complex (Crespo-López et al., 2006). Reactions were performed in group, using the Multiplex PCR Kit (Qiagen). PCR conditions were according to the supplier's recommendation, except for the annealing temperature, which was lowered to 52 °C. The amplification products were analysed with an automated sequencer (ABI 310 Genetic Analyzer).

2.5. Isolation of S. pyrenaicus and S. alburnoides amh

2.5.1. Amplification of amh

To obtain partial sequences of the *S. pyrenaicus* (PP) and *S. alburnoides* (AA) *amh*, representing the two genomes present in hybrids, degenerated primers based on coding sequences of the *amh* gene of *Oryzias latipes* (DQ523689), *Danio rerio* (AY721604) and *Paralichthys olivaceus* (AB166791) were tested, according to the following PCR conditions: pre-heating at 96 °C for 2 min 30 s, 35 cycles at 96 °C for 30 s, 50 °C for 30 s and 72 °C for 1 min 15 s and a final extension at 72 °C for 10 min. The products obtained were sequenced and new sets of specific primers were designed (Table 1).

In order to extend the sequence region of *amh* obtained for *S. alburnoides*, new primers were designed and tested, based on the full-length sequences of *S. pyrenaicus* (Table 1). Samples of non-hybrid (AA) *S. alburnoides* were used as template for amplification.

The forward primers were used in rapid amplification of cDNA ends. First strand cDNA (including the 3'end information of mRNA) was reversed transcribed from testis RNA using a 3′ RACE System (Invitrogen). Amplification was performed using an adaptor primer and the gene specific forward primer. PCR conditions were as follows: pre-heating at 94 °C for 3 min, 30 cycles at 94 °C for 45 s, 55 °C for 25 s and 72 °C for 3 min and a final extension at 72 °C for 15 min. Negative (only using the adaptor primer) and positive (*β-actin*) controls were included in the same reaction. The obtained products were sequenced and alignments were performed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.).

2.5.2. Cloning of amh fragments isolated from hybrid S. alburnoides

To allow the identification of the two genomes (P and A) in hybrids, amplification products were cloned and sequenced separately. Diploid (PA) and triploid (PAA) gonad cDNA samples were used as template for PCR amplification. Products were cloned into PcrII-TOPO vector (Invitrogen), positive colonies were identified by blue/white selection and screened with M13F and M13R universal primers. Positive colonies were randomly picked and sequenced. Sequences were analysed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.) and compared with sequences corresponding to the P and A genomes, previously isolated from *S. pyrenaicus* (PP) and *S. alburnoides* nuclear non-hybrid (AA) males.

2.5.3. Protein alignment and phylogenetic analysis

The deduced amino acid sequence of the *S. pyrenaicus amh* was aligned with *Paralichthys olivaceus* (Pol — AMH, BAD37138), *Anguilla japonica* (Aja — AMH, BAB93107), *Salmo salar* (Ssa — AMH, AAU85130) and *Danio rerio* (Dre — AMH, AY881649) *amh*.

For the same sequences, a phylogenetic tree was constructed using the Neighbour-Joining method as implemented in MEGA version 2.1 (Kumar et al., 2001). For the phylogenetic reconstruction, both amino acid sequences of *amh* from other species and other members of the transforming growth factor-β (TGF-β) superfamily were additionally included: *Homo sapiens* Amh (Hsa — AMH, NP_000470), *Mus musculus* Amh (Mmu — AMH, NP_031471), *Gallus gallus* Amh (Gga — AMH, NP_990361) *Danio rerio* Bmp2a, Bmp2b and Tgfb2 (respectively, Dre — Bmp2a, NP_571434, Dre — Bmp2b, NP_571435 and Dre — TGFbeta2, NP_919366), *Homo sapiens* Bmp2 and Tgfβ2 (Hsa — Bmp2, NP_001191 and Hsa — TGFbeta2, NP_003229). Support for the observed topology was generated by performing 1000 bootstrap replicates.

Table 1
Degenerate primers based on sequences of *Danio rerio, Oryzias latipes* and *Paralichthys olivaceus amh* (position in *D. rerio* sequence is indicated); and specific primers based on the obtained sequences for *S. pyrenaicus* and *S. alburnoides*

Primer	Sequence	Position *D. rerio* seq	Specific for *S. alb/S.pyr*
AMH-F01	5′-ATGCTTTTCCAGACAAGATTTGG-3′	1	
AMH-R01	5′-AGCAGSGWCTCKSTGGAKGACA-3′	986	
AMH-F02	5′-ACACAGWCTGTKTGCATTTC-3′	533	
AMH-R02	5′-ATGTGGSWGTTSAGCAGGA-3′	1508	
AMH-F1 D	5′-GCCTCTGTRMGCTCCT-3′	479	
SqAMH-R1	5′-CATCCACCTATCTCATCAACA-3′		+
SqAMH-F2	5′-TGTTGATGAGATAGGTGGATGG-3′		+
SqAMH-R2	5′-AGGGTCTCGGTGGATGACAC-3′		+
SqAMH-F10	5′-TGTTGCTGACTGTGGCGACT-3′		+
SqAMH-R10	5′-CAGCGACATTCACACTTGGT-3′		+

2.6. Analysis of amh expression in adult gonads of S. pyrenaicus and S. alburnoides

2.6.1. Probe synthesis

Colonies selected after the cloning procedure (representing the P or A genomes) were cultured overnight and purified using E.Z.N.A. Plasmid Miniprep Kit II (PEQLab). Sense and antisense DIG-labelled riboprobes of approximately 650 bp were obtained by digestion with NotI or BamH I and transcription with SP6 and T7 polymerases (DIG RNA Labelling Kit-Roche).

2.6.2. Positive control probe synthesis

In order to obtain a positive control for the *in situ* hybridization protocol on cryosections, the *β-actin* and *vasa* gene orthologs were amplified from *S. pyrenaicus* testis cDNA using primers Act1 5′-CAACGGCTCCGGCATGTG-3′; Act2 5′-TGCCAGGGTACATGGTGG-3′ and Vasa1F 5′-TGTGGACGTGAGTGGCAGCAATC-3′; Vasa1R 5′-TGCTCCACATCACTGCA-3′, according to the following PCR conditions: preheating at 96 °C for 2 min 30 s, 35 cycles at 96 °C for 30 s, 53 °C for 30 s and 72 °C for 1 min 15 s and a final extension at 72 °C for 10 min. The amplified products were cloned into PcrII-TOPO vector (Invitrogen). Insert screening, probe synthesis and purification were performed as described in 2.6.1.

2.6.3. Tissue preparation and sectioning

Testis from adult males of *S. pyrenaicus* and *S. alburnoides* and ovaries of triploid females of *S. alburnoides* were fixed and processed according to Bajanca et al. (2004). Embedded gonads were then frozen and stored at −80 °C. Cryosections of 10 μm were obtained and collected on Superfrost Plus slides (VWR). To better visualize the morphology of adult gonads, some slides were processed for laminin immunhistochemistry (polyclonal rabbit anti-laminin antibody from Sigma) and counterstained with 4′,6-diamidine-2-phenylidole-dihydrochloride (DAPI, 5 μg/ml in PBS 0,1% Triton X-100) as described in Bajanca et al. (2004).

2.6.4. Whole mount in situ hybridisation

In situ hybridisation was performed using a protocol adapted from Jowett (2001). Slides were defrosted at room temperature and hybridization was performed overnight at 65 °C in a humid box (1X SSC/50% formamide). Subsequent washes were performed with 1X SSC/50% formamide (65 °C) and MABT buffer (at 37 °C, rocking). After blocking for one hour with 10% heat-inactivated sheep serum in PBT, slides were incubated overnight in a 1:5000 dilution of anti-digoxigenin antibody. To avoid the effects of possible endogenous phosphatase activity, levamisole was included in both NTMT buffer and the staining solution (BM-purple AP substrate, Roche). Slides were incubated in staining solution overnight and mounted with Aquatex.

2.7. Analysis of amh expression during S. alburnoides development

2.7.1. Experimental crosses

Selected specimens of *S. pyrenaicus* and *S. alburnoides* were used to perform four experimental crosses. Females were striped and their eggs exposed to sperm on a Petri dish. Gametes were gently mixed with a brush and water was added. Fertilized eggs were laid on a mesh and incubated in the aquarium.

2.7.2. Collection of embryos

Embryos from the four experimental crosses were collected at different stages of development. For RNA extraction embryos were directly frozen at −80 °C. Embryos to be used in the whole mount *in situ* hybridization protocol were fixed overnight in 4% paraformaldehyde in PBS at 4 °C, followed by dehydration with a methanol series and stored in 100% methanol at −20 °C. For the *in situ* hybridization protocol on cryosections, embryos were fixed as described in 2.6.3 and stored at −80 °C. Embryos that died during the time period of 50 days were collected for DNA extraction and genotyping and preserved in ethanol at 4 °C. Parents and progeny were genotyped using the same set of microsatellite loci.

2.7.3. In situ hybridisation

Hybridisation was conducted on whole embryos following a protocol adapted from Henrique et al. (1995), using *amh* probes and *β-actin* probes as controls. Embryos were dehydrated in an ethanol series and included in Tecnovit 8100 and dried for subsequent sectioning in a LKB ultra microtome.

3. Results and discussion

3.1. Genotyping of adult specimens and progeny of experimental crosses

Individuals were identified according to morphological characters and divided in three groups: *S. pyrenaicus*, hybrid *S. alburnoides* of both sexes and *S. alburnoides* nuclear non-hybrid males. The genotypes of individuals of the two species (*S. pyrenaicus* and *S. alburnoides*) from which gonads were collected were determined based on the information of three loci (LCO3, LCO4 and LCO5) (Table 2). The alleles were attributed to each genome (A or P) according to the scoring already developed for Southern populations of the *S. alburnoides* complex (Crespo-López et al., 2006). The same approach was used to identify individual genotypes of parents and progeny of experimental crosses (Table 3). As evidenced by multilocus genotyping, all experimental crosses yielded diploid progeny: Crosses I and III resulted in all AA progeny, while Crosses II and IV resulted in PA offspring, as expected. According to the reproductive dynamics of the complex (for review see Alves et al., 2001) and the progeny genotype, only specimens from two crosses (I and III) were used in the subsequent expression analysis. According to what has been previously shown through exhaustive analysis of experimental crosses (Alves et al., 1998), one should expect that progeny resulting from these crosses would be all male (AA genotype), while the genotype from crosses II and IV should include both males and females.

Presently, there is no molecular marker in *S. alburnoides* that would allow the distinction of sexes during early developmental stages, so we only characterized the *amh* expression pattern in

Table 2
Summary of amplification results for the three microsatellite loci (LCO3, LCO4, LCO5): Species (*S. alburnoides* and *S. pyrenaicus*); Sex; *n* (number of the individual); *P* (ploidy); alleles identified for each locus; and (G) genotype

Species	Sex	P	n	LCO3	LCO4	LCO5	G
S. pyr	M	2n	131	$243^{P}243^{P}$	$238\ ^{P}238^{P}$	$137^{P}143^{P}$	PP
S. pyr	M	2n	135	$243^{P}243^{P}$	$242\ ^{P}242^{P}$	$137^{P}143^{P}$	PP
S. pyr	M	2n	137	$243^{P}243^{P}$	$240^{P}240^{P}$	$137^{P}137^{P}$	PP
S. pyr	F	2n	5	$243^{P}243^{P}$	$238\ ^{P}240^{P}$	$137^{P}137^{P}$	PP
S. alb	M	2n	104	$247^{A}247^{A}$	$262^{A}280^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	106	$247^{A}247^{A}$	$286^{A}302^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	107	$247^{A}247^{A}$	$268^{A}270^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	110	$247^{A}247^{A}$	$278^{A}280^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	85	$243^{P}247^{A}$	$238^{P}284^{A}$	$131^{A}143^{P}$	PA
S. alb	F	2n	86	$243^{P}247^{A}$	$240^{P}290^{A}$	$131^{A}137^{P}$	PA
S. alb	F	3n	89	$239^{A}243^{P}247^{A}$	$272^{P}296^{A}296^{A}$	$131\ ^{A}131^{A}143^{P}$	PAA
S. alb	F	3n	90	$243^{P}247^{A}249^{A}$	$268^{P}276^{A}276^{A}$	$131^{A}131^{A}143^{P}$	PAA
S. alb	F	3n	117	$243^{P}247^{A}247^{A}$	$238^{P}282^{A}302^{A}$	$131^{A}131^{A}137^{P}$	PAA

individuals that we could unambiguously attribute to either sex. To avoid confusing results and to increase the chances of actually determining the role of *amh* in early *S. alburnoides* sex determination (early onset of *amh* expression in female larvae has only been reported in one fish species) (Klüver et al., 2007), only AA progeny were used in the analysis of gene expression during development.

3.2. Isolation of amh

Amplification with different *amh* primer combinations resulted in products composed of several sequence segments of the *amh* gene coding sequence of *S. pyrenaicus* and *S. alburnoides*. The 3′ RACE experiment using Squal-F2 primer resulted in the amplification of the full-length coding sequence of *S. pyrenaicus* (P genome).

A 1618 bp product was obtained, corresponding to the full coding sequence of the *amh* gene of *S. pyrenaicus* (EU136185). The full coding sequence of *S. alburnoides amh* could not be obtained (only a 1411 bp fragment including part of the TGF-β domain (EU136186)) .The alignment of the two sequences revealed an overall identity of 95% between the genes of the two species, which would be expected by their phylogenetic relateness. An overall conservation of the TGF-β and the AMH specific domains, which characterize the family, was observed for the *S. alburnoides* and *S. pyrenaicus* Amh, particularly in comparison with zebrafish (Fig. 1). The conservation of functional motifs, in comparison to other Amh proteins (Fig. 1) and the phylogenetic proximity to the Amh group of teleosts, revealed by the analysis involving different proteins of the TGF-β superfamily (Fig. 2) confirmed the isolation of the *amh* ortholog in *S. pyrenaicus* and *S. alburnoides*.

Table 3
Summary of the results from the four experimental crosses: (*C*) cross; (*S. alburnoides* and *S. pyrenaicusi*) species; (*P*) sex and ploidy; (*N*) number of offspring with the same allelic combination; identified alleles for each locus (alleles in brackets were inferred) and (*G*) genotype

C	Species	P	N	LCO3	LCO4	LCO5	G
I	*S.alb* (26)	♀3n		$243\ ^{P}247\ ^{A}$	$238^{P}\ 266^{A}\ 272\ ^{A}$	$131\ ^{A}\ 137\ ^{P}$	PAA
	S.alb (27)	♂2n		$247^{A}\ (247\ ^{A})$	$274^{A}\ 286^{A}$	$131\ ^{A}\ (131^{A})$	AA
	Progeny						
	S.alb	(2n)	3	$247^{A}\ (247\ ^{A})$	$266^{A}286^{A}$	$(131^{A})\ 131^{A}$	AA
			4	$247^{A}\ (247\ ^{A})$	$266^{A}274^{A}$	$(131^{A})\ 131^{A}$	AA
			1	$247^{A}\ (247\ ^{A})$	$272\ ^{A}286^{A}$	$(131^{A})\ 131^{A}$	AA
			1	$247\ ^{A}\ (247\ ^{A})$	$272\ ^{A}286^{A}$	$(131^{A})\ 131^{A}$	AA
II	*S.pyr*(28)	♀2n		$243\ ^{P}\ (243\ ^{P})$	$240^{P}\ (240^{P})$	$137\ ^{P}\ 143\ ^{P}$	PP
	S.alb (29)	♂2n		$247^{A}\ (247\ ^{A})$	$268^{A}\ 280^{A}$	$131^{A}\ (131^{A})$	AA
	Progeny						
	S.alb	(2n)	5	$243\ ^{P}\ 247\ ^{A}$	$240^{P}\ 268^{A}$	$131^{A}\ 143\ ^{P}$	PA
			6	$243\ ^{P}\ 247\ ^{A}$	$240^{P}280^{A}$	$131^{A}\ 143\ ^{P}$	PA
III	*S.alb* (38)	♀3n		$243\ ^{P}\ 247\ ^{A}$	$-\ 270\ ^{A}$	$131^{A}\ 137^{P}$	PAA
	S.alb (39)	♂2n		$247^{A}\ (247\ ^{A})$	$280^{A}\ 306^{A}$	$131^{A}137^{P}$	AA
	Progeny						
	S.alb	(2n)	1	$247^{A}\ (247\ ^{A})$	$270^{A}\ 280^{A}$	$131^{A}\ (131^{A})$	AA
	S.alb	(2n)	4	$247^{A}\ (247\ ^{A})$	$270^{A}\ 280^{A}$	$131^{A}\ (131^{A})$	AA
IV	*S.alb* (44)	♀3n		$243\ ^{P}\ 247\ ^{A}(247^{A})$	$240^{P}\ 270^{A}\ 282^{A}$	$131^{A}(131^{A})137^{P}$	PAA
	S.pyr(41)	♂2n		$243\ ^{P}\ (243\ ^{P})$	–	$137\ ^{P}(137^{P})$	PP
	Progeny						
	S.alb	(2n)	1	$243\ ^{P}\ 247\ ^{A}$	$240^{P}\ 270^{A}$	$131^{A}\ 137^{P}$	PA
	S.alb	(2n)	1	$243\ ^{P}\ 247\ ^{A}$	$240^{P}\ 282^{A}$	$131^{A}\ 137^{P}$	PA

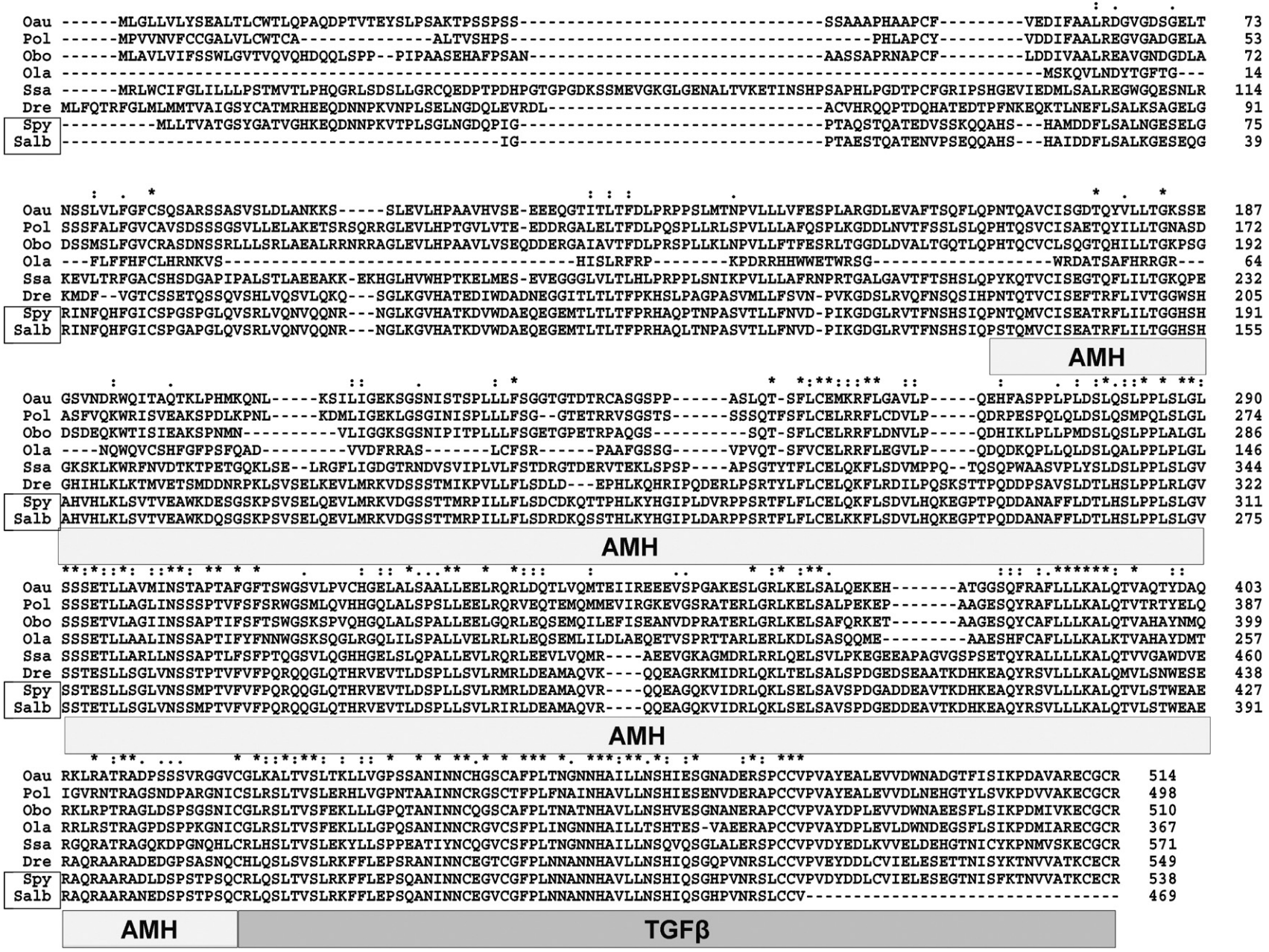


Fig. 1. Protein alignment of *S. pyrenaicus* AMH with other related members of the AMH family: (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved-substitution. Conserved AMH and TGFβ domains that characterize the family are highlighted in grey.

3.3. Expression of amh in the adult gonads is restricted to specific cell lineages

In the adult male gonads of the bisexual maternal ancestor *S. pyrenaicus* (PP) *amh* expression is confined to specific cellular locations, namely to cells surrounding the germ cell lineage (Fig. 3a). A morphological comparison with the gonad organization of other cyprinid species (Parenti and Grier, 2004) and the localization of the seminiferous tubule basement membrane through laminin immunohistochemistry (Fig. 3) suggests that the positive signal obtained with the *amh* probe corresponds to expression in Sertoli cells (Fig. 4a). There is a clear similarity between the observed patterns of *amh* expression and the ones described recently for zebrafish (Rodriguez-Mari et al., 2005), thus implying an overall conservation of the *amh* expression dynamics in the two cyprinid species. Accordingly, *amh* is also expressed in the adult ovaries of *S. pyrenaicus* and *S. alburnoide*s, more specifically in follicular cells surrounding the primordial and primary oocytes (Fig. 4d). This non-dimorphic adult expression and the specific localisation have been described not only for cyprinid species but also in the medaka (Klüver et al., 2007) and in mammals (Lasala et al., 2004). Despite our initial hypothesis of possible disruption of gene expression patterns in the hybrids, no significant differences were observed in *amh* expression pattern in *S. alburnoides* AA and PA (Fig. 4b and c, respectively) male gonads when compared to *S. pyrenaicus* (Fig. 4a). In *Drosophila* hybrid species, misregulation of male-specific genes seems to be correlated to hybrid sterility (Michalak and Noor, 2004). In *S. alburnoides*, on the other hand, normal gene expression patterns seem to be maintained, which is in accordance with the fertility of diploid hybrid gonads (revised in Alves et al., 2001; Crespo-López et al., 2006).

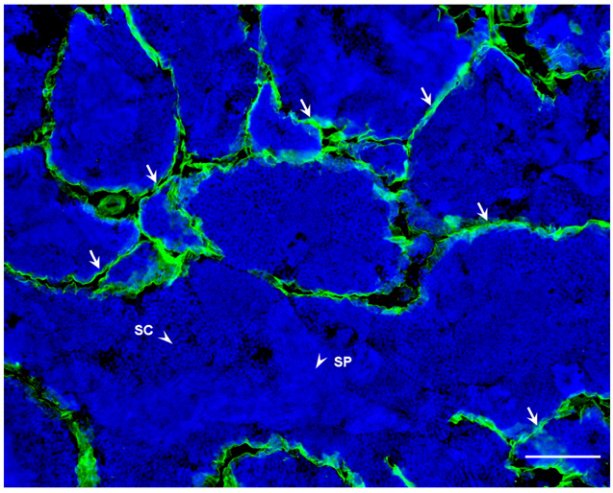


Fig. 3. Immunoreactivity for laminin (green) in adult testis of *S. pyrenaicus* (PP genotype). The laminin-containing basement membrane (arrows) surrounds the seminiferous tubules; arrowheads indicate different stages of germ cell differentiation: spermatocytes (SC) spermatozoa (SP); nuclei are stained with DAPI (blue). Scale bar = 50 μm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Early onset amh expression during male development

In the all male AA genotype embryos, *amh* expression was observed as early as 4 days after hatching and expression

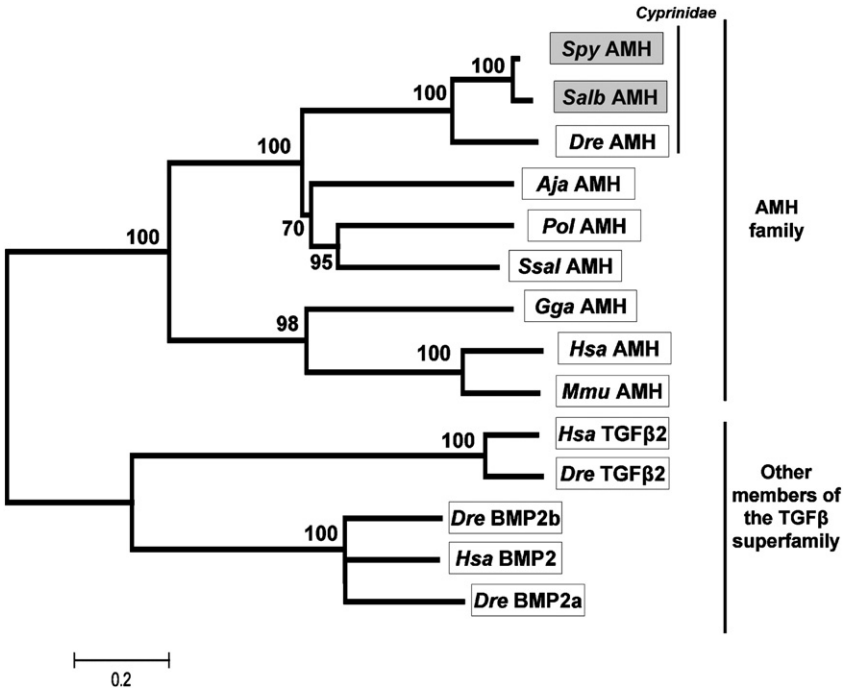


Fig. 2. Unrooted Neighbour-Joining tree based on Poison corrected model obtained among the Amh family (Spy — AMH-*Squalius pyrenaicus* Pol — AMH-*Paralichthys olivaceus*, Aja — AMH-*Anguilla japonica*, Ssa — AMH-*Salmo salar*, Dre — AMH-*Danio rerio*, Gga — AMH-*Gallus gallus*, Mmu — AMH-*Mus musculus*; Has — AMH-*Homo sapiens*) and other protein families belonging to the TGF-beta superfamily (Dre—Bmp2a, Dre—Bmp2b, Dre—TGFbeta2—*Danio rerio* BMP2a, BMP2b and TGFB2; Hsa—Bmp2, Hsa—TGFbeta2 — *Homo sapiens* BMP2 and TGFB2). Bootstrap values are shown above the branches.

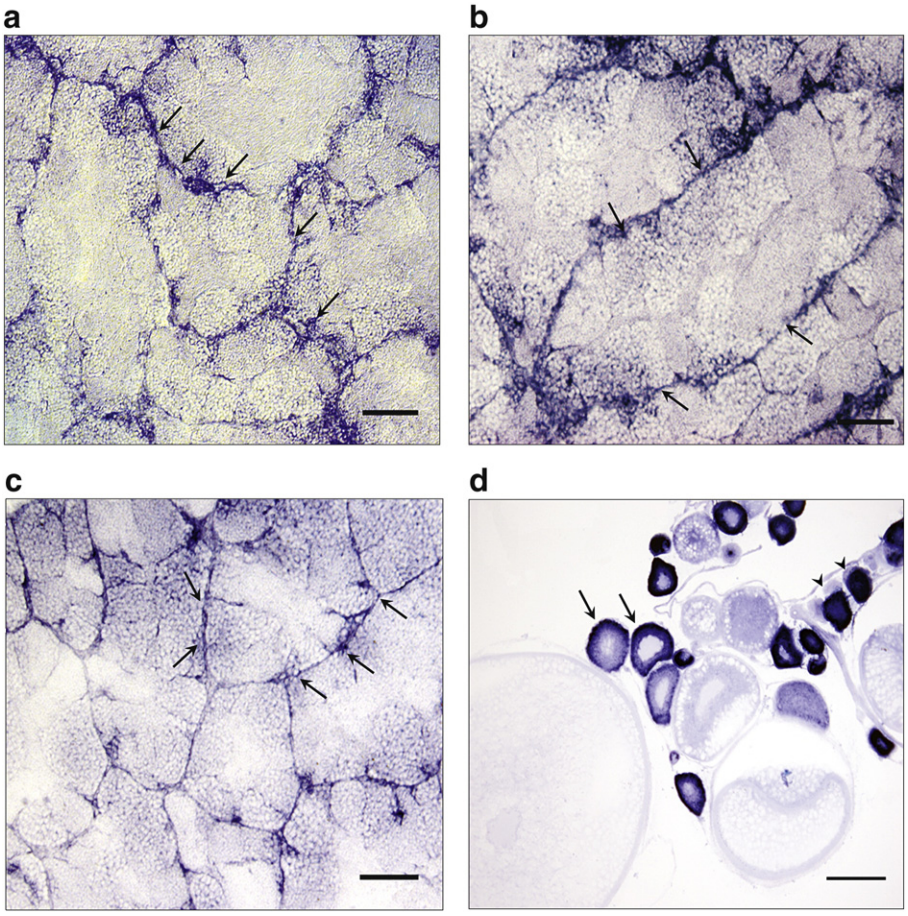


Fig. 4. Expression pattern of *amh* in the adult gonads *S. pyrenaicus* and *S. alburnoides*: (a) *S. pyrenaicus* testis (PP genotype); (b) *S. alburnoides* testis (AA genotype); and (c) *S. alburnoides* testis (PA genotype). Arrows indicate *amh* positive signals and presumptive location of Sertoli cells. Scale bar=100 μm; (d) *S. alburnoides* ovary (PAA genotype). Positive *amh* signals in follicular cells surrounding oocyes are indicated by arrows (primary oocytes) and arrowheads (primordial oocytes). Scale bar=50 μm.

remains detectable until 34 days after hatching (Fig. 5). *Vasa* expression, as a marker of the PCG lineage (Knaut et al., 2000) was detected at all developmental stages analysed, thus confirming the location of the developing gonads. RT-PCR results (data not shown) revealed *amh* expression at several stages between 4 and 34 dah, thus suggesting a role in early development of the male gonad and in adult testes.

The present results, combined with the ones obtained for adult gonads, suggest that the pathway involving *amh* expression and its role in male gonad development, as described for model fish species (Rodriguez-Mari et al., 2005; Klüver et al., 2007) is globally unaltered in *S. alburnoides*.

4. Conclusions

We isolated, cloned and characterized the *amh* orthololog in *S. alburnoides* and *S. pyrenaicus*. This study is the first one to address the molecular biology of sex determination during development in this complex, thus constituting a stepping-stone towards the elucidation of which molecular pathways underlie the complex system of the correlation between sex and genotype in *S. alburnoides*. No significant differences were observed at the sequence or expression patterns in the male adult gonads between the maternal species *S. pyrenaicus* and the hybrids. The observation of an overall conservation of both sequence and expression characteristics could also be extended through comparison with other cyprinid species, such as zebrafish. Similarly to what is observed in zebrafish, *amh* expression is confined to the Sertoli cell lineage in male gonads of *S. alburnoides*, and to the follicular cells in the adult female gonad. The analysis of *amh* expression during development confirmed that it is expressed early during male gonad development in *S. alburnoides* suggesting that, in analogy to the situation in other teleosts, it could play an important role in the development of the male gonad.

Our results point towards the conservation of the role of *amh* in early male gonad development and in adult gonad physiology in both sexes in the hybrid genetic context of *S. alburnoides*. However, the strong correlation between genotype and sex in the *S. alburnoides* complex, as suggested by the occurrence of an all-male lineage, implies that sex determination should have a strong genetic basis. It is possible that other upstream elements of the sex determination cascade could be contributing to the sex distribution phenomenon and that *amh* would maintain its unaltered role and contribute to male gonad development as in non-hybrid cyprinid species (Wang and Orban, 2007). It is also possible that the location and timing of expression of sex determination genes would remain unaltered in *S. alburnoides*

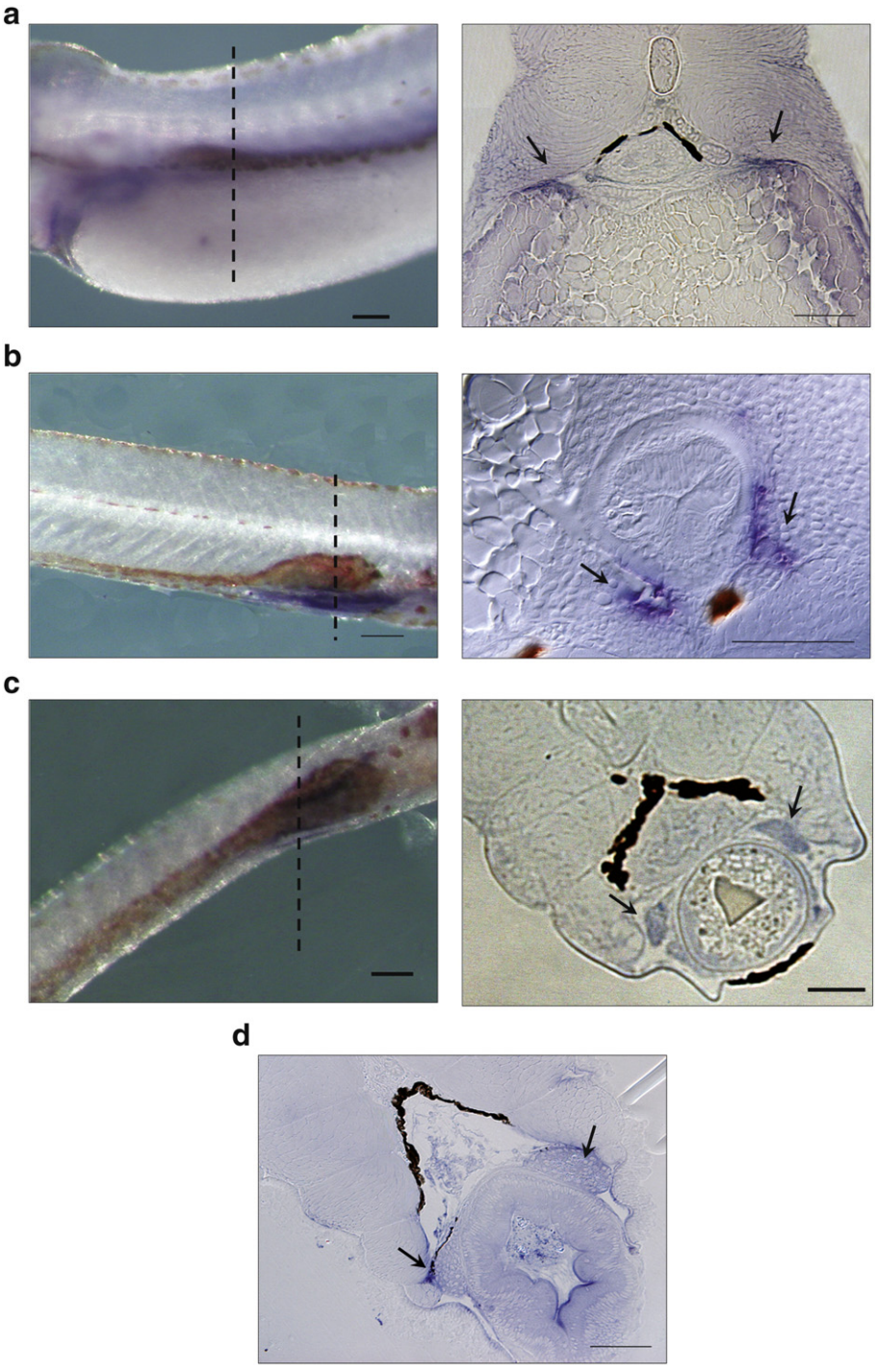


Fig. 5. Expression pattern of *amh* in the developing gonad of *S. alburnoides*: (a) In a whole embryo and section (dashed line) at 4 days after hatching (dah); (b) In whole embryo and section (dashed line) at 8 dah; (c) In the whole embryo and section (dashed line) at 14 dah; (d) Section at 34 dah. Arrows indicate positive signals in the location of developing gonads. Scale bar = 100 μm.

hybrids, but that slight variations in gene dosage levels during development, as a result of the combination of one or more copies of two different genomes, could lead to deviations in the normal sex determination process and, consequently, a deviation in sex ratios. In fact, sex determination cascades in higher vertebrates have been shown to be quite sensitive to dosage of individual genes (Parker et al., 1999). It is possible that regulatory variations of gene expression would occur in early developmental stages, according to genome contribution. Only when a doubtless identification of sexes in hybrid (PA) individuals can be performed will this possibility be fully addressed. By following this candidate gene approach, we ultimately aim to enlighten particular aspects of *S. alburnoides* sex distribution and gather additional evidence that would contribute to the elucidation of the evolution of sex determination mechanisms, within the context of hybridisation.

Acknowledgements

The authors thank Carina Cunha for the help in the field and collaboration in the breeding experiments and maintenance of the fish. This research was supported by a grant from FCT to I. Pala (SFRH/ BD/ 19286/ 2004), by the DFG-Graduiertenkolleg 1048 (Molecular Basis of Organ Development in Vertebrates) through a PhD fellowship to N.K. and by the projects POCTI/ BSE/40868 (FEDER participation) and PTDC/BIA-BDE/ 65586/2006.

References

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 1998. Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. Mol. Biol. Evol. 15, 1233–1242.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 2001. Evolution in action through hybridization and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111, 375–385.

Alves, M.J., Collares-Pereira, M.J., Dowling, T.E., Coelho, M.M., 2002. The genetics of maintenance of an all-male lineage in the *Squalius alburnoides* complex. J. Fish Biol. 60, 649–662.

Bajanca, F., Luz, M., Duxson, M.J., Thorsteinsdóttir, S., 2004. Integrins in the mouse myotome: developmental changes and differences between the epaxial and hypaxial lineage. Dev. Dyn. 231, 402–415.

Barroiller, J.F., Guigen, Y., Forstier, A., 1999. Endocrine and environmental aspects of sex differentiation if fish. Cell. Mol. Life Sci. 55, 910–931.

Crespo-López, M.E., Duarte, T., Dowling, T.E., Coelho, M.M., 2006. Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in Tejo and Guadiana rivers: an approach with microsatellites. Zoology 109, 277–286.

Dawley, R.M., Goddard, K.A., 1988. Diploid-triploid mosaics among unisexual hybrids of the minnows *Phoxinus neogaeus*. Evolution 42, 649–659.

Durlinger, A.L., Visser, J.A., Themmen, A.P., 2002. Regulation of ovarian function: the role of anti-Müllerian hormone. Reproduction 124, 601–609.

Halm, S., Rocha, A., Miura, T., Prat, F., Zanuy, S., 2007. Anti-Mullerian hormone (*AMH*/AMH) in the European sea bass: Its gene structure, regulatory elements, and the expression of alternatively-spliced isoforms. Gene 388, 148–158.

Henrique, D., Adam, J., Myat, A., Chitnis, A., Lewis, J., Ish-Horowicz, D., 1995. Expression of a Delta homologue in prospective neurons in the chick. Nature 375, 787–790.

Josso, N., Clemente, N., 2003. Transduction pathway of anti-Mullerian hormone, a sex-specific member of the TGF-beta family. Trends Endocrinol. Metab. 14, 91–97.

Josso, N., Racine, C., di Clemente, N., Rey, R., Xavier, F., 1998. The role of anti-Mullerian hormone in gonadal development. Mol. Cell. Endocrinol. 145, 3–7.

Jowett, T., 2001. Double *in Situ* Hybridisation Techniques in Zebrafish. Methods 23, 345–358.

Klüver, N., et al., 2007. Differential expression of anti-Müllerian hormone (*amh*) and anti-Müllerian hormone receptor type II (*amhrII*) in the teleost Medaka. Dev. Dyn. 236, 271–281.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17, 1244–1245.

Lasala, C., Carre-Eusebe, D., Picard, J.Y., Rey, R., 2004. Subcellular and molecular mechanisms regulating anti-Mullerian hormone gene expression in mammalian and nonmammalian species. DNA Cell Biol. 23, 572–585.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., Sakaizumi, M., 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417, 559–563.

Michalak, P., Noor, M.A.F., 2004. Association of mis-expression with sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. J. Mol. Evol. 59, 277–282.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Miura, T., Miura, C., Konda, Y., Yamauchi, K., 2002. Spermatogenesis-preventing substance in Japanese eel. Development 129, 2689–2697.

Nanda, I., et al., 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc. Natl. Acad. Sci. U. S. A. 99, 11778–11783.

Knaut, H., Pelegri, F., Bohmann, K., Schwarz, H., Nüsslein-Volhard, C., 2000. Zebrafish vasa RNA but not its protein is a component of the germ plasm and segregates asymmetrically before germline specification. J. Cell Biol. 149, 875–888.

Parenti, L.R., Grier, H.J., 2004. Evolution and Phylogeny of Gonad Morphology in Bony Fishes. Integr. Comp. Biol. 44, 333–348.

Parker, K.L., Schedi, A., Schimmer, B.P., 1999. Gene interactions in gonadal development. Annu. Rev. Physiol. 61, 417–433.

Rodriguez-Mari, A., Yan, Y.L., Bremiller, R.A., Wilson, C., Canestro, C., Postlethwait, J.H., 2005. Characterization and expression pattern of zebrafish Anti-Mullerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. Gene Expr. Patterns 5, 655–667.

Wang, X.G., Orban, L., 2007. Anti-Müllerian hormone and 11 β-hydroxilase show reciprocal expression to that of aromatase in the transforming gonad of zebrafish males. Dev. Dyn. 236, 1329–1338.

Wilhelm, D., Palmer, S., Koopman, P., 2007. Sex determination and gonadal development in mammals. Physiol. Res. 87, 1–28.

Yoshinaga, N., Shiraishi, E., Yamamoto, T., Iguchi, T., Abe, S., Kitano, T., 2004. Sexually dimorphic expression of a teleost homologue of Mullerian inhibiting substance during gonadal sex differentiation in Japanese flounder, Paralichthys olivaceus. Biochem. Biophys. Res. Commun. 322, 508–513.

CHAPTER 3

Candidate genes for sex determination in the *Squalius alburnoides* complex: an initial characterization of sex cascade elements in the context of a hybrid genome

Irene Pala, Manfred Schartl, Sólveig Thorsteinsdóttir, Maria Manuela Coelho

In prep

Candidate genes for sex determination in the *Squalius alburnoides* complex: an initial characterization of sex cascade elements in the context of a hybrid genome

Irene Pala [a], Manfred Schartl [b], Sólveig Thorsteinsdóttir [a, c], Maria Manuela Coelho [a]

[a] Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

[b] Physiologische Chemie I, Biozentrum, University of Würzburg, Würzburg, Germany

[c] Instituto Gulbenkian de Ciência, Oeiras, Portugal

Abstract

When addressing the question of sex determination in functional and evolutionary perspectives, teleost fishes constitute a very interesting study basis, due to the broad diversity of mechanisms exhibited within the group. In model fish species, several genes have been shown to participate in sex determination cascades and a number of mechanisms have been proposed to account for the process. Difficulties arise when transposing these questions to a "non-model context". In hybrid fish, the knowledge is even scarcer and the gene interactions underlying the process of sex determination have not been identified. This is the case of the cyprinids of the *Squalius alburnoides* complex, in which hybridisation and polyploidy have contributed to the emergence of diverse reproductive mechanisms, altering both oogenesis and spermatogenesis, affecting sex ratios and resulting in strong correlations between sex and genotype. We have initiated the isolation and characterization of regulatory elements of the cascade potentially involved in sex determination in *S. alburnoides* (*dmrt1*, *wt1*, *dax1* and *figla*) and analysed their expression during development and in adult gonads by *in situ* hybridisation. In the adults, an overall conservation in the cellular location of the transcripts of these genes was observed between the hybrid *S. alburnoides* and the bisexual parental species *S. pyrenaicus* and for most genes, the expression patterns also suggested the maintenance of roles in the adult gonad, comparatively to other teleost species. Some novel interesting features have also emerged, such as the contribution of *dmrt1* in the adult ovaries and the non-dimorphic expression of *figla*, a reported ovarian marker in gonads of both sexes in *S. alburnoides* and *S. pyrenaicus*. The potential contribution of each gene to the sex determination process was assessed based on the timing and location of expression. *Dmrt1* and *wt1* transcripts were found at early stages of male development in *S. alburnoides* and are most likely implicated in the process of gonad development. By establishing a working basis of different genes, the present work constitutes the necessary initial step towards the study of the genetic basis of the complex process of sex determination in such an unorthodox hybrid system.

Keywords: cyprinids; hybridisation; sex determination; gonad development

1. Introduction

The genetic basis of sex determination and gonad development processes has been intensively studied in several animal groups and strict regulatory networks have been revealed in birds and mammals (Manolakou *et al.*, 2006). For most fish, however, the mystery remains unsolved. In this group, a wide spectrum of sexual systems is represented, ranging from hermaphroditism to gonochorism and a frequent switch between mechanisms as diverse as environmental or strictly genetic based sex determination can be observed (Baroiller, 1999; Devlin and Nagahama, 2002; Schartl, 2004). Even in model species such as the zebrafish *Danio rerio*, the mechanism underlying sex determination is largely unknown and a candidate gene approach has been conducted over the years in an attempt to elucidate which players could be involved in the process (Trant *et al.*, 2001; Von Hofsten *et al.*, 2005; Rodriguez-Mari *et al.*, 2005; Jørgensen *et al.*, 2008). Such information is far scarcer in non-model species and the high diversity of processes of sex determination that characterize the teleost fish impairs any possibility of mechanism inference based on evolutionary proximity. Among non-model fish, some particular groups offer an additional challenge, as they present reproductive and sex distribution alterations that imply an even broader diversity of regulatory processes (Otto & Whitton, 2000). Such is the case of the object of our study: the *Squalius alburnoides* complex of hybrid fish. This complex is endemic from the Iberian Peninsula and resulted from interspecific hybridisation between females of *Squalius pyrenaicus* (P genome) and males of an unknown species related to *Anaecypris hispanica* (A genome). The complex includes individuals of various ploidies that intercross through highly diverse reproductive modes that range from clonal inheritance to normal meiosis, hybridogenesis or meiotic hybridogenesis (in which one genome is excluded from gamete formation) (for a review see Alves *et al.*, 2001). Sex ratio distribution is clearly altered in the *S. alburnoides* complex, with a strong bias towards triploid females of PAA genotype. The complex also includes a fertile lineage of individuals of AA genotype, apparently composed exclusively by males (Alves *et al.*, 2001). The strong correlation between genomic constitution and sex is an interesting feature to explore in terms of the sex determination process in these fish. Until now, no sex chromosomes were identified that would indicate a chromosomal based sex determination process in this complex, although chromosome heteromorphism suggesting a ZWfemale/ ZZmale system has been reported for the maternal ancestor *S. pyrenaicus* (Collares-Pereira *et al.* 1998). Data from experimental crosses revealed that such a system could not fully explain sex determination in *S. alburnoides* and that in addition to female determinants on the W chromosome, a minimum of one non-W-linked gene would have to be expressed differently in hybrid and nonhybrid genome combinations to account for the results obtained. Moreover, the

"strength" of these sex determining factors might be variable depending on populations and parental species (Alves *et al.* 1998).

Thus, and in the absence of reports of environmental clues or hormonal induced responses playing a role in the regulation of sex determination in the *S. alburnoides* complex, all evidence points towards a gene-mediated mechanism. So far, only one sex determination gene (the anti-Müllerian hormone, *amh*) has been isolated in *S. alburnoides*, and its expression characterized both during development and in the adult (Pala *et al.*, 2008a). To fully explore the possibility of a genetic based sex determination mechanism, an initial set of possible candidates has to be established. Then it has to be determined whether those potential players in the sex determination cascade are expressed in *S. alburnoides* at critical stages of gonadal differentiation and whether sex specific expression patterns could imply additional roles in the adult.

The *Dmrt1* (*doublesex* and *mab-3* related transcription factor 1) gene is part of a family of genes that share a common zinc-finger DNA binding motif (the DM domain). The DM domain was initially identified in genes that occupy key positions in sex determination pathways of the fruit fly *Drosophila melanogaster* (the *doublesex* gene) (Erdman & Burtis, 1993) and in the nematode *Caenorhabditis elegans* (*mab-3* gene) (Raymond *et al.*, 1998), but *Dmrt1* homologues have been isolated, shown to be expressed mainly in the adult and developing gonads, and correlated to sex determination cascades in a number of species (reviewed in Ferguson-Smith, 2007). In humans, although acting downstream in the sex determination cascade, *DMRT1* has been implicated in some types of XY sex reversal (reviewed in Veitia *et al.*, 2001). In birds, its location in the Z chromosome (Nanda *et al.*, 1999) and the higher dosage in males (Raymond *et al.* 1999a; Smith *et al.* 1999a; Shan *et al.* 2000) make it a very good candidate for male sex determining gene. Also, *Dmrt1* expression leads to sex reversal in chicken embryos, which apparently confirms its important role in the male determination and testis differentiation pathways (Smith *et al.*, 2003). In fish, it has been shown to be involved in male sex determination pathways in the Nile tilapia *Oreochromis niloticus* (Guan *et al.*, 2000) and in the rainbow trout *Oncorhynchus mykiss* (Marchand *et al.*, 2000), and to contribute both to male and female gonad development in zebrafish (Guo *et al.*, 2005). Most prominently, in the medaka *Oryzias latipes*, a *dmrt1* duplicate in the Y chromosome (*dmrt1bY*/ *DMY*) (Nanda *et al.*, 2002; Matsuda *et al.* 2002) has been identified as the master male sex determination gene in this species.

The Wilm's tumor suppressor gene (*Wt1*) is a key regulator of urogenital development. The gene encodes a nuclear protein containing four zinc fingers, acting both as a transcription factor and in RNA processing. In mammals, at least 16 different protein isoforms of *Wt1* can be generated from alternative splicing, alternative translation initiation and RNA editing (reviewed in Englert, 1998). In non-mammalian vertebrates two isoforms occur, resulting from alternative splicing (Kent *et al.*,

1995). The most important functional difference within all isoforms results from the insertion of three additional aminoacids (KTS) by splicing events: the WT1 (-KTS) isoform acts as a transcription factor, while WT1 (+KTS) is mainly involved in RNA processing (Niksic *et al.*, 2004; Bor *et al.*, 2006). In the medaka, an additional splice variant of Wt1, *wt1a*_dE4, with differences in exon 4 is present in both +KTS and –KTS isoforms (Klüver *et al.*, 2008).

Evidence of *Wt1* roles during development comes from different animal species. In mammals, *WT1* mutations can result in Wilm's tumor, Denys-Drash and Frasier syndromes, affecting the development of gonads and genitalia (reviewed in Rivera and Haber, 2005). The involvement of *Wt1* in the differentiation of urogenital structures and the maintenance/proliferation of the cells of the bipotential gonad has also been shown in other mammals (Armstrong *et al.*, 1993), in the chicken (Smith *et al.*, 1999b), in reptiles (Western *et al.*, 2000) and amphibians (Semba *et al.*, 1996). *Wt1* orthologs have been isolated in telosts species, such as the medaka *O. latipes* (Fedorova *et al.*, 2008; Klüver *et al.*, 2008), zebrafish *D. rerio* (Bollig *et al.*, 2006), the Japanese eel *Anguilla japonica* (Nakatsuru *et al.*, 2000) and the rainbow trout *O. mykiss*, (Brunelli *et al.*, 2001). Considering its role in gonad differentiation, a few possible transcriptional targets of *Wt1* have been identified, such as *Dax1*, *Sf1* and *Gata 4* (Kim *et al.*, 1999; Wilhelm and Englert, 2002).

In contrast to *dmrt1* and *wt1,* the other candidate addressed in this work, *dax1*, was traditionally considered as an "anti-testis", ovarian determination gene. It is a member of the nuclear receptors superfamily, with a ligand-binding domain at the C-terminal and LxxLL motifs, which enable the binding and interaction with potential targets (reviewed in Iyer & McCabe, 2004). Dax1 acts as a transcriptional repressor of several genes involved in the development and steroidogenic activity of adrenal and gonadal structures, such as aromatase *cyp19* (Wang *et al.*, 2001), anti-müllerian hormone *amh* (Tremblay & Viger, 2001), estrogen and androgen receptors (Zhang *et al.*, 2000; Holter *et al.*, 2002), and most notably was shown to inhibit Sf1-mediated transcriptional transactivation (Clipsham *et al.*, 2003). *Dax1* expression is normally restricted to specific tissues, namely the genital ridge and largely coexpressed with *Sf1*, antagonizing its cooperation with WT1 and GATA-4 in the regulation of AMH (Nachtigal *et al.*, 1998; Tremblay & Viger, 2001). In humans, *DAX1* duplication results in dosage sensitive sex reversal (DSS) syndrome, causing male to female sex reversals in individuals with normal *SRY* expression (Bardoni *et al.*, 1994). Also, mutations in this gene often result in disruptions of normal gonad development. Mouse models have elucidated a role of *Dax1* in normal testicular development and male fertility, whereas in females the knockout of the gene mainly affected endocrine function, but apparently not normal ovarian development (Yu *et al.*, 1998). These evidences supported a role as "anti-testis" factor only under certain conditions, and contradicted the classical view of Dax1 as an ovarian-determining gene.

Due to its relevant role in the hypothalamic-pituitary-adrenal-gonadal (HPAG) axis and gonad development as well as the interaction with sex steroid producing enzymes or receptors in gonad differentiation, *Dax1* orthologs have been isolated and studied in different groups such as birds (Smith *et al.*, 2000), reptiles (Western *et al.*, 2000; Maldonado *et al.*, 2002) and amphibians (Sugita *et al.*, 2001). In fish, *dax1* expression patterns have been investigated both regarding its function in adrenal development (in the zebrafish, *D. rerio*) (Zhao *et al.*, 2006) and its contribution to gonad development and differentiation in several species, such as the nile tilapia *O. niloticus* (Wang *et al.*, 2002), the European sea bass *Dicentrarchus labrax* (Martins *et al.*, 2007) and the medaka *O. latipes* (Nakamoto *et al.*, 2007).

The fourth gene that is reported to be involved in early sex differentiation of several groups is the Factor in the germline alpha (*Figla*), a germ cell specific basic helix-loop-helix transcription factor. This factor plays a key role in folliculogenesis, by coordinating the regulation of the zona pellucida (zpc) genes (Liang *et al.*, 1997). Absence of *Figla* expression has been shown to affect female, but not male fertility in mice, suggesting a critical role in female germline development (Soyal *et al.*, 2000).

In a system such as the *S. alburnoides* complex, in which no genomic information is available, our strategy was to isolate potential candidates for sex determination and assess, based on their conservation, functional characteristics and expression patterns, whether they could have a role in the establishment of the phenotypic sex. This analysis creates an initial framework of possible genes that can then be further investigated. In the present work we report the isolation of the *dmrt1*, *wt1*, *dax1* and *figla Squalius* orthologs, the characterization of their specific features with respect to potential functionality, and the description of their expression in the gonads of the bisexual species and hybrids by *in situ* hybridisation. Finally, we have evaluated the potential for a more extensive analysis of their contribution for male or female gonad development in the *S. alburnoides* complex.

2. Materials and Methods

2.1. Samples

The specimens used in this study were previously collected from the Tejo River Basin (River Raia), as described in Pala *et al.*, 2008a. A total of 9 gonad samples, 3 *of S. pyrenaicus* and 6 of *S. alburnoides* (2 nuclear non-hybrid, 2 diploid, and 2 triploid hybrids) were selected from the global sample already analysed and genotyped using microsatellites (for details see Pala *et al.*, 2008a). An additional *S. pyrenaicus* sample was included and processed as described in Pala *et al.*, 2008a. The selected samples included individuals of different genotype, sex and ploidy level in order to comprise the

different forms that compose the complex and represent the two genomes present in its Southern distribution: the P and A genomes (a summary of the selected samples is presented in Table I-a).

Table I. Summary of amplification results for the three microsatellite loci (LCO3, LCO4, LCO5) in adult samples (a) and in the parents and progeny of the experimental cross (b): Species (*S. alburnoides* and *S. pyrenaicus*); Sex; P (ploidy); n (number of the individual); nr (number of offspring with the same allelic combination); alleles identified for each locus; and (G) genotype.

a- Adult samples							
Species	**Sex**	**P**	**n**	**LCO3**	**LCO4**	**LCO5**	**G**
S.pyr	M	2n	131	$243^{P}243^{P}$	$238^{P}238^{P}$	$137^{P}143^{P}$	PP
S.pyr	M	2n	137	$243^{P}243^{P}$	$240^{P}240^{P}$	$137^{P}137^{P}$	PP
S.pyr	F	2n	132	$243^{P}243^{P}$	$240^{P}240^{P}$	$137^{P}143^{P}$	PP
S.pyr	F	2n	5	$243^{P}243^{P}$	$238^{P}240^{P}$	$137^{P}137^{P}$	PP
S.alb	M	2n	106	$247^{A}247^{A}$	$286^{A}302^{A}$	$131^{A}131^{A}$	AA
S.alb	M	2n	110	$247^{A}247^{A}$	$278^{A}280^{A}$	$131^{A}131^{A}$	AA
S.alb	M	2n	85	$243^{P}247^{A}$	$238^{P}284^{A}$	$131^{A}143^{P}$	PA
S.alb	F	2n	86	$243^{P}247^{A}$	$240^{P}290^{A}$	$131^{A}137^{P}$	PA
S.alb	F	3n	90	$243^{P}247^{A}249^{A}$	$268^{P}276^{A}276^{A}$	$131^{A}131^{A}143^{P}$	PAA
S.alb	F	3n	117	$243^{P}247^{A}247^{A}$	$238^{P}282^{A}302^{A}$	$131^{A}131^{A}137^{P}$	PAA
b- Embryos- Experimental cross							
Parents							
Species	**Sex**	**P**		**LCO3**	**LCO4**	**LCO5**	**G**
S.alb	F	3n		$243^{P}247^{A}$	$238^{P}266^{A}272^{A}$	$131^{A}137^{P}$	PAA
S.alb	M	2n		$247^{A}247^{A}$	$274^{A}286^{A}$	$131^{A}131^{A}$	AA
Progeny			**nr**				
S.alb		2n	3	$247^{A}247^{A}$	$266^{A}286^{A}$	$131^{A}131^{A}$	AA
S.alb		2n	4	$247^{A}247^{A}$	$266^{A}274^{A}$	$131^{A}131^{A}$	AA
S.alb		2n	1	$247^{A}247^{A}$	$272^{A}286^{A}$	$131^{A}131^{A}$	AA
S.alb		2n	1	$247^{A}247^{A}$	$272^{A}286^{A}$	$131^{A}131^{A}$	AA

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted from adult gonads of *S. pyrenaicus* and *S. alburnoides* using the TRIZOL reagent (Gibco-BRL) according to the supplier's recommendation. First strand cDNA was synthesized with RevertAid™ First Strand cDNA Synthesis Kit (Fermentas), using random hexamer primer.

2.3. Isolation of *S. pyrenaicus* and *S. alburnoides* candidate genes

2.3.1. Amplification of *dmrt1, wt1, dax1* and *figla*

A partial fragment of the *dmrt1* gene of *S. pyrenaicus* (PP) and *S. alburnoides* (AA) was amplified using degenerate and specific primers (Table II), according to the conditions described in Pala *et al.*,

2008b. Products corresponding to the A and P genomes were sequenced and analysed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.).

To obtain initial sequence information for the isolation of the *wt1* gene, degenerate primers (Table II – WT1 F1a; WT1 R1a; WT1 F1b; WT1 R1b) were designed based on coding sequences of this gene from *D. rerio* (NM_131046; NM_001039634), *A. japonica* (AB030741) and *O. mykiss* (AF334670; AF334671). In species such as *D. rerio* two *wt1* copies have been reported (Bollig *et al.*, 2006), and the primer pairs referred to above were designed taking into account the specific sequence differences of each of the two orthologues - a and b.

cDNA samples of *S. pyrenaicus* (PP) and *S. alburnoides* (AA) gonads were used as templates for amplification according to the following PCR conditions: pre-heating at 96°C for 2 min 30 s, 35 cycles at 96°C for 30 s, 52°C for 45 s and 72°C for 1 min 15 s and a final extension at 72°C for 10 min.

Amplification was only successful in *S. pyrenaicus* and partial sequences for both primer pairs were obtained. In order to extent the length of the obtained product, an additional forward primer was designed (Table II). Amplification was performed using this primer and primers WT1 R1a; R1b, according to the conditions described above and an annealing temperature of 54°C. Sequences were obtained, analysed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.) and their homology with other *wt1* transcripts was confirmed through BLAST comparison.

Amplification of a partial sequence of the *dax1* gene of *S. pyrenaicus* and *S. alburnoides* was performed using primers designed based on the zebrafish *dax1* (ENSDART00000020212) sequence (Table II). cDNA samples of *S. pyrenaicus* (PP) and *S. alburnoides* (AA and PAA) gonads were used as templates for amplification, according to the following PCR conditions: pre-heating at 96°C for 2 min 30 s, 35 cycles at 96°C for 30 s, 56°C for 45 s and 72°C for 1 min 15 s and a final extension at 72°C for 10 min. Products were sequenced, analysed and alignments were performed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.).

Primers for the *Figla* gene were designed based on zebrafish sequences (NM_198919) (Table II) and tested using *S. pyrenaicus* (PP – male and female) and *S. alburnoides* (AA- male, PAA- female) gonad samples as templates. Products were amplified according to the conditions described above, sequenced and analysed.

2.3.2. Cloning

Amplification products of the four genes were cloned into pDrive Cloning Vector (Qiagen), positive colonies were identified by blue/white selection and screened with M13F and M13R universal primers. Positive colonies were randomly picked and sequenced. Sequences were analysed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.) and compared with sequences corresponding to

the P and A genomes, previously isolated from *S. pyrenaicus* (PP) and *S. alburnoides* nuclear non-hybrid (AA) males.

Table II. Primer sequences (primers designed in the present work), references and GenBank accession numbers for each gene.

Gene	Primer	Sequence/ Ref	Accessionnr.
dmrt1	DMRT1- F1 D	Pala *et al.*, 2008b	EU199439
	DMRT1 zf-R3		EU199440
wt1	WT1 F1a	5'- ATGGGYTCWGAYGTKCGTGACC-3'	
	WT1 R1a	5'- GCMTAYCCTGGCTGCAACAAA -3'	
	WT1 F1b	5'- AGCTGCAGCACHCAGTAYC-3'	FJ587497
	WT1 R1b	5'- TTGSTCAKRTTKCTCTGRTGC-3'	
	WT1 F5	5'- CACTTCTCYGGACAGYTCA -3'	
dax1	DAX1 F1	5'- TCTGCGAGGATGGCTTACTT -3'	FJ587498
	DAX1 R1	5'- CATGGAGAGAGCGAGGAAGA -3'	FJ587499
figla	FIG F2	5'- CGAGATAAAGCTGTGAAGAGG-3'	FJ587500
	FIG R1	5'- CCTGACGTCATTGTGACCAG -3'	FJ587501

2.3.3. Protein alignment and phylogenetic analysis

The deduced amino acid sequence of the *S. pyrenaicus* (PP-DMRT1) and *S. alburnoides* (AA-DMRT1) *dmrt1* was aligned with *D. rerio* (Dre-DMRT1, AAQ04555) *Takifugu rubripes* (Tru-DMRT1, BAE16952), *O. mykiss* (Omy-DMRT1, AAG17544) and *O. latipes* (Ola-DMRT1, AAL02165). The deduced amino acid sequence of the *S. pyrenaicus* (PP-WT1) *wt1* ortholog was aligned with *D. rerio* (Dre-a-WT1, NP_571121; Dre-b-WT1, NP_001034723), *O. mykiss* (Omy-a-WT1, AAK52719; Omy-b-WT1, AAK52721), *A. japonica* (Aja-WT1, BAA90558) and *O. latipes* (Ola-WT1, BAC10628). Alignments were also performed for the deduced amino acid sequences of the two remaining genes. For *dax1*, aminoacid sequences of *D. rerio* (Dre-DAX, Q1L693), *Dicentrarchus labrax* (Dla-DAX1, CAG17628), *O. latipes* (Ola-DAX1, NP_001104259) and *O. niloticus* (Ony-DAX1, AAN17672) were aligned with the *putative S. pyrenaicus* (PP-DAX1) and *S. alburnoides* (AA-DAX1) orthologs. The deduced amino acid sequence of *Figla* of *S. pyrenaicus* (PP_FiglaM and PP_FiglaF – from male and female samples) and *S. alburnoides* (from male samples AA- FiglaM) were aligned with *D. rerio* (Dre-FIGLA, NP_944601), *Tetraodon nigroviridis* (Tni-FIGLA, ACH91670) and *Kryptolebias marmoratus* (Kma-FIGLA, ABG89136). Specific protein domains included in the deduced amino acid sequences for the Iberian species were identified by homology search using Simple Modular Architecture Research Tool (SMART) (Letunic *et al.*, 2008).

Phylogenetic trees were constructed using the Neighbour-Joining method as implemented in MEGA version 2.1 (Tamura *et al.*, 2007).

For the phylogenetic reconstruction, both amino acid sequences of DMRT1 from other species and other proteins sharing the DM DNA-binding domain were additionally included: *Homo sapiens* DMRT1 (Hsa-DMRT1, CAB99335), *Gallus gallus* DMRT1 (Gga-DMRT1, Q9PTQ7), *H. sapiens* DMRT2 (DMRT2-Hsa, AL358976), *O. latipes* DMRT2 (DMRT2-Ola, AAL02163), *H. sapiens* DMRT3 (Hsa-DMRT3, CAB99336), *T. rubripes* DMRT3 (DMRT3-Tru, BAE16954) and *D. rerio* DMRT3 (DMRT3-Dre, AAU89440). For WT1, in addition to the sequences used in the protein alignments, amino acid sequences from orthologs from other species were also used: *H. sapiens* WT1 (Hsa-WT1, CAI95759), *G. gallus* WT1 (Gga-WT1, NP_990547), *Mus musculus* WT1 (Mmu-WT1, P22561) and *Rattus norvegicus* WT1 (Rno-WT1, NP_113722). For DAX1, phylogenetic reconstruction was based on the teleost and other species' DAX1 proteins as well as other protein family members, including: *H. sapiens* DAX1 (Hsa-DAX1, AAC13875), *M. musculus* DAX1 (Mmu-DAX1, NP_031456), *G. gallus* (Gga-DAX1, NP_989924) and *O. niloticus* DAX2 (Oni-DAX2, ABB88832). For FIGLA, phylogenetic relationships were inferred based on the teleost sequences already used in the alignment and FIGLA orthologs from other species: *H. sapiens* FIGLA (Hsa- FIGLA, AAS48452), *R. norvegicus* FIGLA (Rno-FIGLA, XP_575589) and *Xenopus (Silurana) tropicalis* FIGLA (Xtr FIGLA, NP_001016342).

Support values for the observed topologies were generated by 1000 bootstrap replicates.

2.3.4. Protein structure comparison

Secondary structure was predicted, based on the amino acid sequences of the *dmrt1* gene fragments of *S. pyrenaicus* and *S. alburnoides* using the PSIPRED method (Jones, 1999) as implemented in PSIPRED server (Bryson *et al.*, 2005). Tertiary structure prediction of the *Squalius* Dmrt1 proteins was performed by homology modelling, using ESyPredED (Lambert et al, 2002).

2.4. Analysis of candidate gene expression in adult gonads of *S. pyrenaicus* and *S. alburnoides*

2.4.1. Probe synthesis

Colonies selected after the cloning procedure (representing the P or A genomes) were cultured overnight and purified using the Miniprep Kit (Invitrogen). Sense and antisense DIG-labelled riboprobes were obtained by digestion with *Hind*III or *Bam*HI (in the case of *dmrt1*, *wt1* and *figla*), *Hind*III and *Kpn*I (*dax1*) and transcription with SP6 and T7 polymerases (DIG RNA Labelling Kit - Roche).

2.4.2. Tissue preparation and sectioning and *in situ* hybridisation

Testis from adult males of *S. pyrenaicus* and *S. alburnoides* (AA and PA) and ovaries of diploid (PA) and triploid (PAA) females of *S. alburnoides* were fixed and processed according to Bajanca *et al.* (2004). Embedded gonads were then frozen and stored at -80°C. Cryosections of 10µm were obtained and collected on Superfrost Plus slides (VWR).

In situ hybridisation was performed as described in Pala *et al.* (2008a).

2.5. Analysis of candidate gene expression during *S. alburnoides* development

2.5.1. Collection of embryos and *in situ* hybridisation

Embryos at different developmental stages and of specific genotype (diploid AA) were selected from the progeny of the experimental Cross I (see Pala *et al.*, 2008a for the cross and Table I- b for information about the selected progeny). In the absence of a molecular sex marker to unambiguously identify sexes in the *S. alburnoides* complex we relied on the extremely strong correlation between AA genotype and male phenotype, with only one report of a single female individual of AA genotype in 20 years of intensive study of the complex (Sousa-Santos *et al.*, 2006). Thus, we selected a group of samples for which we could determine sex (in this case male sex) with a high degree of confidence. Progeny had been fixed, processed, frozen and stored at -80°C according to Bajanca *et al.* (2004). Cryosections of 14µm were obtained and collected on Superfrost Plus slides (VWR).

In situ hybridisation was performed as described in Pala *et al.* (2008a). Briefly, slides were defrosted at room temperature and hybridisation was performed overnight at 65°C in a humid box (1X SSC/ 50% formamide). Subsequent washes were performed with 1X SSC/ 50% formamide (65°C) and MABT buffer (at 37°C, rocking). After blocking for one hour with 10% heat-inactivated sheep serum in PBT, slides were incubated overnight in a 1:5000 dilution of anti-digoxigenin antibody. To avoid the effects of possible endogenous phosphatase activity, levamisole was included in both NTMT buffer and the staining solution (BM-purple AP substrate, Roche). Slides were incubated in staining solution overnight and mounted with Aquatex.

3. Results

3.1. Isolation of candidate genes – structural and phylogenetic analysis

3.1.1. *dmrt1*

Amplification of a partial fragment of the *dmrt1* gene was performed using primers DMRT1- F1 D and DMRT1 zf-R3 and sequences of 463bp (from *S. pyrenaicus* -PP) and 457bp (from *S. alburnoides*-AA) were obtained (Table II). Both fragments included the specific DM DNA binding domain that characterizes this class of proteins (Fig.1). The conservation of functional motifs, in comparison to other DMRT1 proteins (Fig. 1) and the phylogenetic proximity to the Dmrt1 genes of other teleosts, revealed by the analysis involving different proteins that share the conserved DM domain (Fig. 2) confirmed the isolation of the Dmrt1 ortholog in *S. pyrenaicus* and *S. alburnoides*.

A two amino acid (Threonine and Aspargine) deletion followed by an amino acid substitution (Leucine by Isoleucine) was observed on the predicted sequence of *S. alburnoides* (A genome) compared to *S. pyrenaicus* (P genome), but no relevant differences in the secondary structure of the protein were found as a consequence of the genome-specific polymorphism (Fig. 3). The modelling method used for tertiary structure prediction only provided results for the sequence between amino acid 13 and 61, which mainly corresponds to the conserved DM domain and is outside the polymorphic area between *S. alburnoides* and *S. pyrenaicus*. The impossibility of predicting structure outside this domain results from the fact that the method applied here (as most methods available and tested by us) is based on homology search and thus is limited by identity with structures for which modelling has already been established.

3.1.2. *wt1*

Of all primer combinations tested, only the ones involving WT1-F1a and, WT1-R1a and WT1-R1b, respectively (Table II) resulted in the production of sequence segments that could be identified as *wt1.* BLAST comparison revealed an overall similarity of all amplified fragments with WT1a transcripts of other teleosts. A larger PCR fragment of 786 bp was amplified with primers WT1 F5 and WT1-R1a (Table II), using *S. pyrenaicus* gonad cDNA as template. The closer homology of our isolate to a-form Wt1 proteins of other telelosts was confirmed both in the phylogenetic tree comparing other Wt1 orthologs (Fig. 4) and the alignments of the predicted amino acid sequence (Fig. 5). Three distinct domains were identified in the *S. pyrenaicus* Wt1 isolate (Fig. 5): a Wilm's tumour specific domain, and two (of the four) zinc finger DNA-binding motifs that characterize the C terminus of this transcription factor.

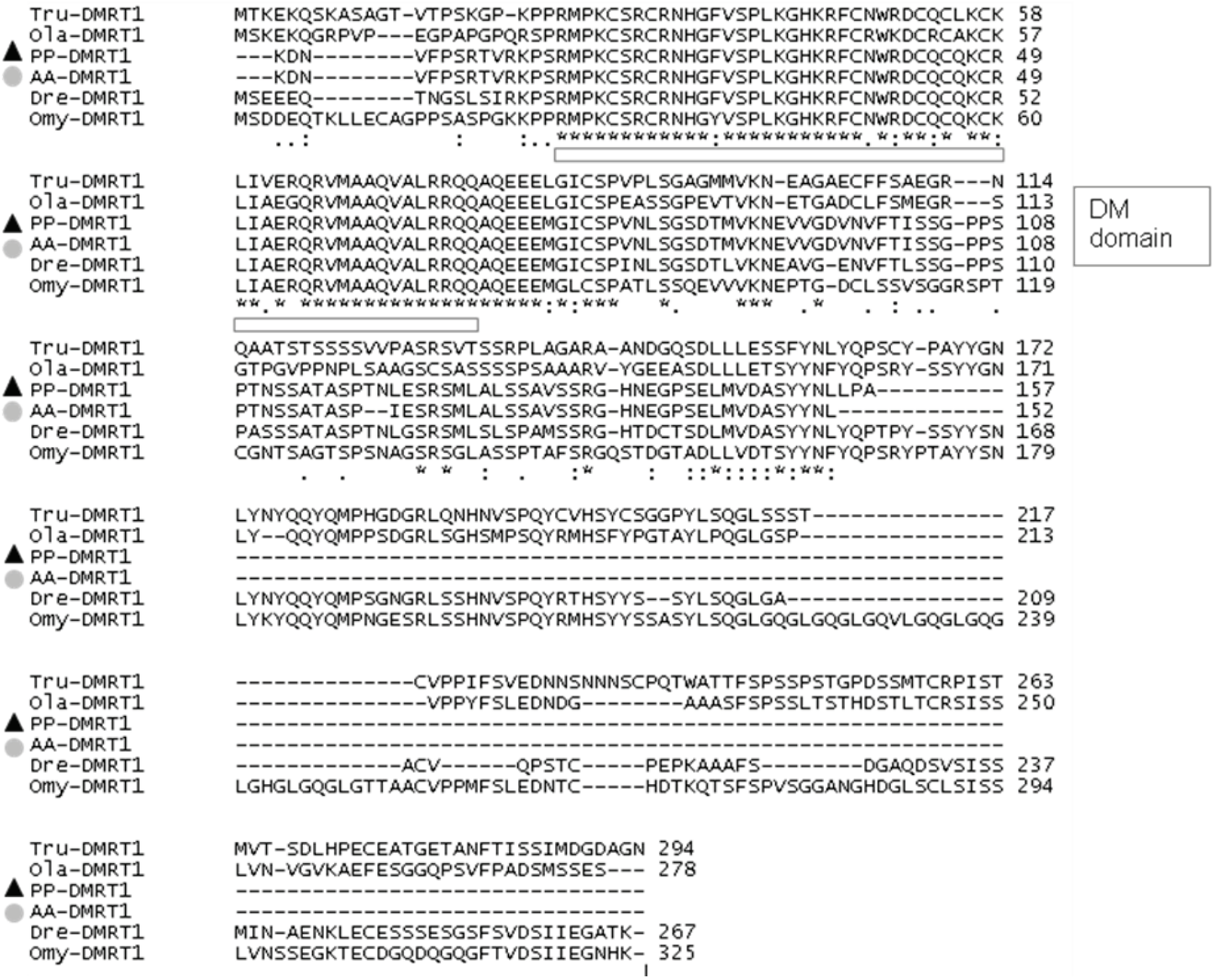


Figure 1. Protein alignment of *Squalius* Dmrt1 with Dmrt1 orthologs of other teleosts: (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved substitution. The conserved DM domain that characterizes the family is highlighted in grey.

▲ *S. pyrenaicus* (PP) ; ● *S. alburnoides* (AA)

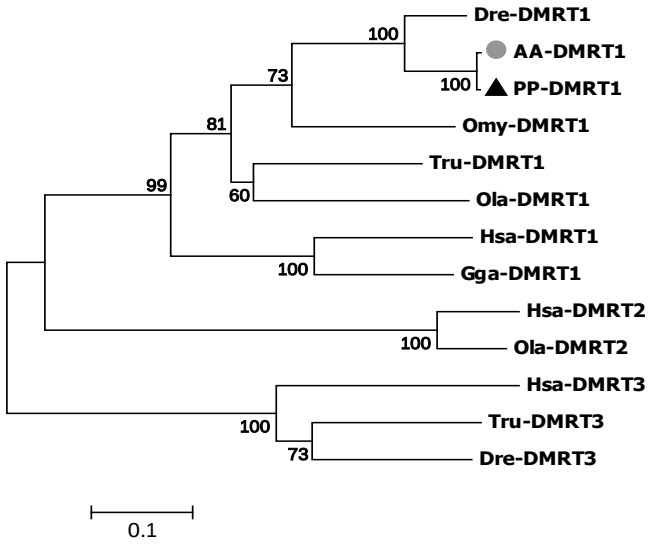


Figure 2. Unrooted Neighbour-Joining tree obtained among the DMRT1 family (Dre-DMRT1- *Danio rerio*, Tru-DMRT1- *Takifugu rubripes*, Omy-DMRT1- *Oncorhynchus mykiss*, Ola-DMRT1- *Oryzias latipes,* Hsa-DMRT1- *Homo sapiens*, Gga-DMRT1- *Gallus gallus*) and other proteins that share the conserved DM domain (DMRT2-Hsa- *H. sapiens,* DMRT2-Ola- *O. latipes,* Hsa-DMRT3- *H. sapiens,* DMRT3-Tru- *T. rubripes,* DMRT3-Dre- *D. rerio*). Bootstrap values are shown above the branches. ▲ *S. pyrenaicus* (PP) ; ● *S. alburnoides* (AA)

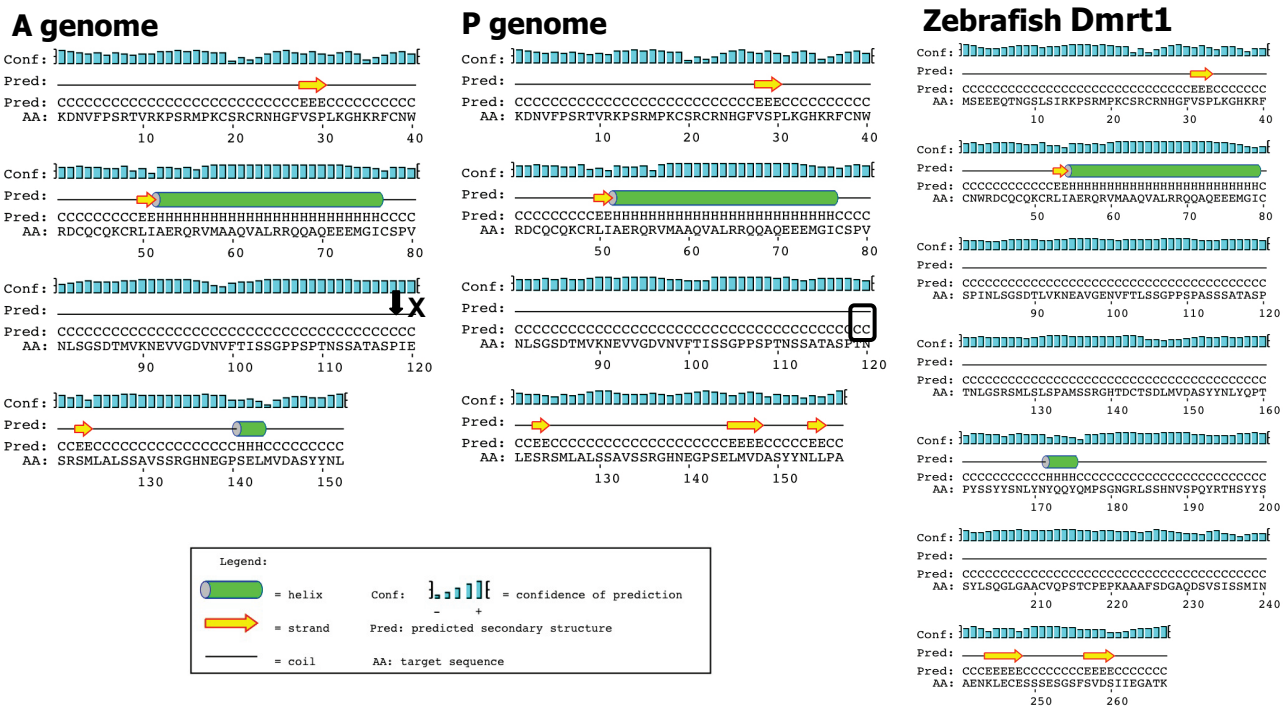


Figure 3. Secondary structure prediction for the partial Dmrt1 proteins of *S. pyrenaicus* (P genome), *S. alburnoides* (A genome) and the full amino acid sequence of zebrafish Dmrt1. The location of the two amino acid deletion is indicated in the A genome (arrow) and the corresponding site is highlighted in the P genome (square).

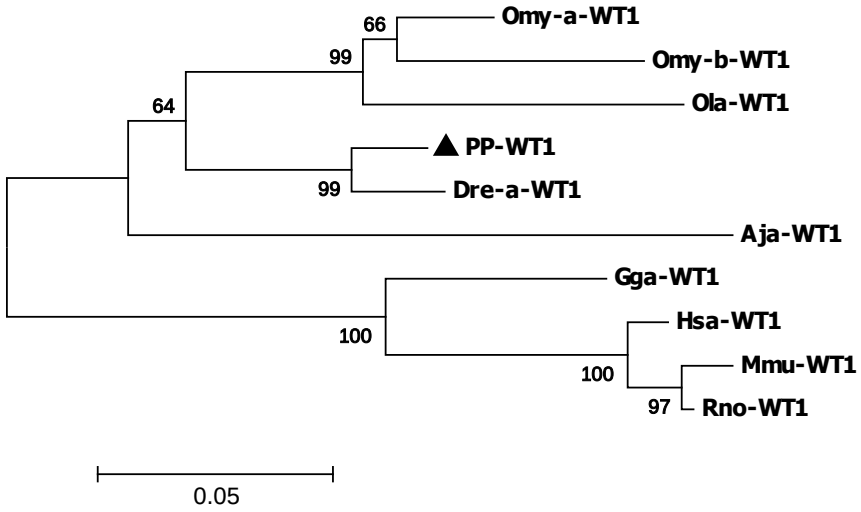


Figure 4. Unrooted Neighbour-Joining tree obtained among the teleost Wt1 proteins (Dre-a-WT1- *Danio rerio*, Omy-a-WT1; Omy-b-WT1- *Oncorhynchus mykiss*, Aja-WT1- *Anguilla japonica* and Ola-WT1- *Oryzias latipes*) and other WT1 orthologs (Hsa-WT1- *Homo sapiens*, Gga-WT1- *Gallus gallus*, Mmu-WT1- *Mus musculus*, and Rno-WT1- *Rattus norvegicus*. Bootstrap values are shown above the branches. ▲ *S. pyrenaicus* (PP)

3.1.3. *dax1*

Fragments of approximately 680bp, corresponding to a partial coding sequence of the *dax1* gene were amplified with primers DAX1-F1 and DAX1-R1 using gonad cDNA samples of *S. pyrenaicus* and *S. alburnoides* as templates (Table II). Alignment of the predicted amino acid sequences revealed 96% identity with Dax1 protein of zebrafish and 60-64% identity with Dax1 proteins of other teleosts (Fig. 6). The phylogenetic reconstruction (Fig. 7) revealed a close evolutionary proximity to the Dax1 protein of zebrafish, with both cyprinids grouping with the Dax1 group of proteins of other teleosts.

The domain prediction analysis confirmed that the *Squalius* isolates included functional regions that characterize Dax1 proteins and that are related to their role as transcription factors (Fig. 6). The *Squalius dax1* gene encodes for two LxxLL motifs, an important feature for mediating protein-protein interactions and a typical characteristic of non-mammalian Dax1 proteins (as opposed to the four LxxLL motifs in mammals) (Zhao *et al.*, 2006). Furthermore, a conserved ligand-binding domain, characteristic of hormone receptors was predicted within our isolate. Thus, although it corresponds to a partial coding sequence, the *dax1* fragment isolated from *S. alburnoides* and *S. pyrenaicus* includes all relevant features that characterize this family of proteins.

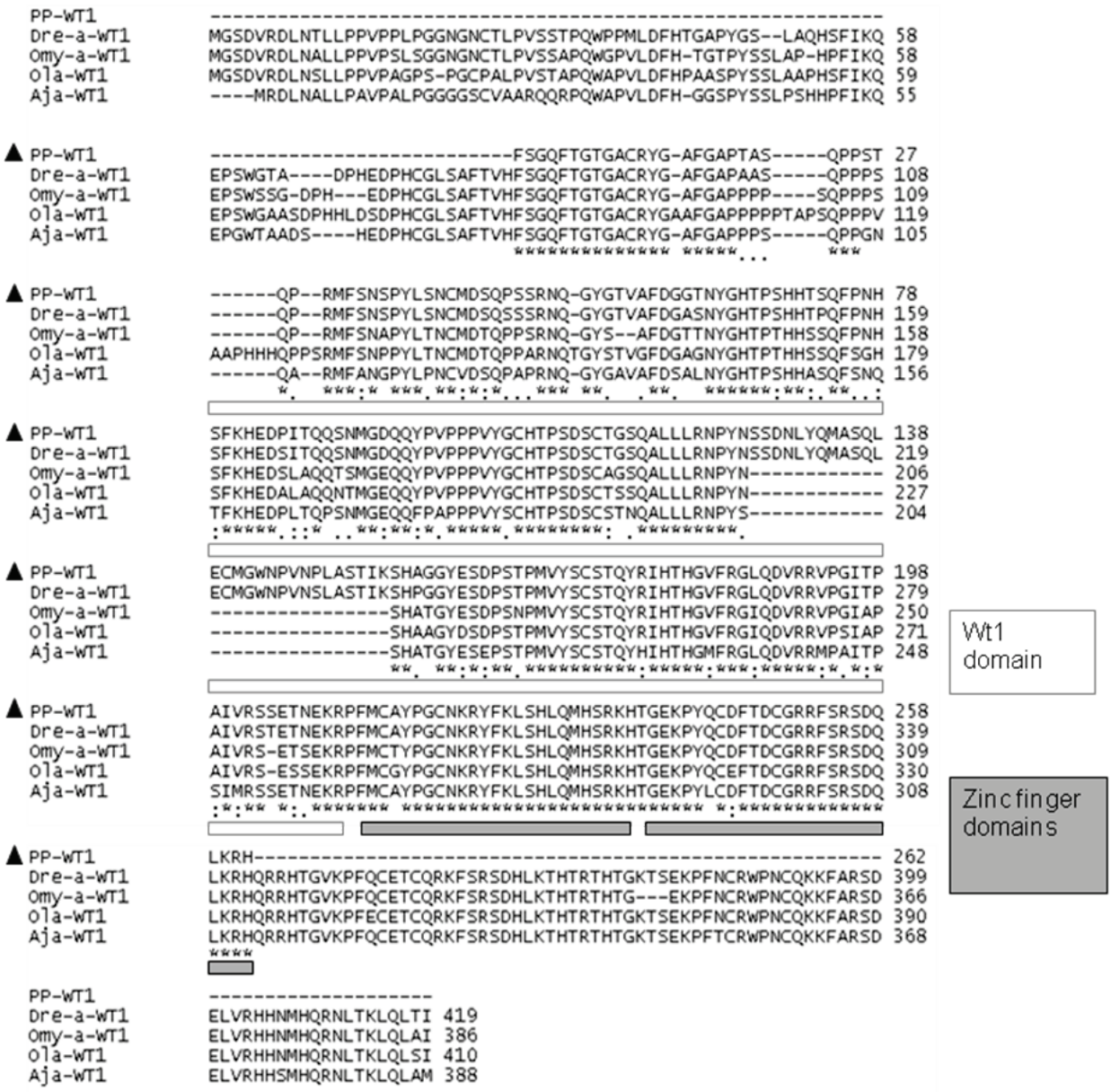


Figure 5. Protein alignment of *S. pyrenaicus* Wt1 with Wt1 orthologs of other teleosts: (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved-substitution. The zinc finger and WT1 specific domains that characterize the protein are highlighted in grey. ▲ *S. pyrenaicus* (PP)

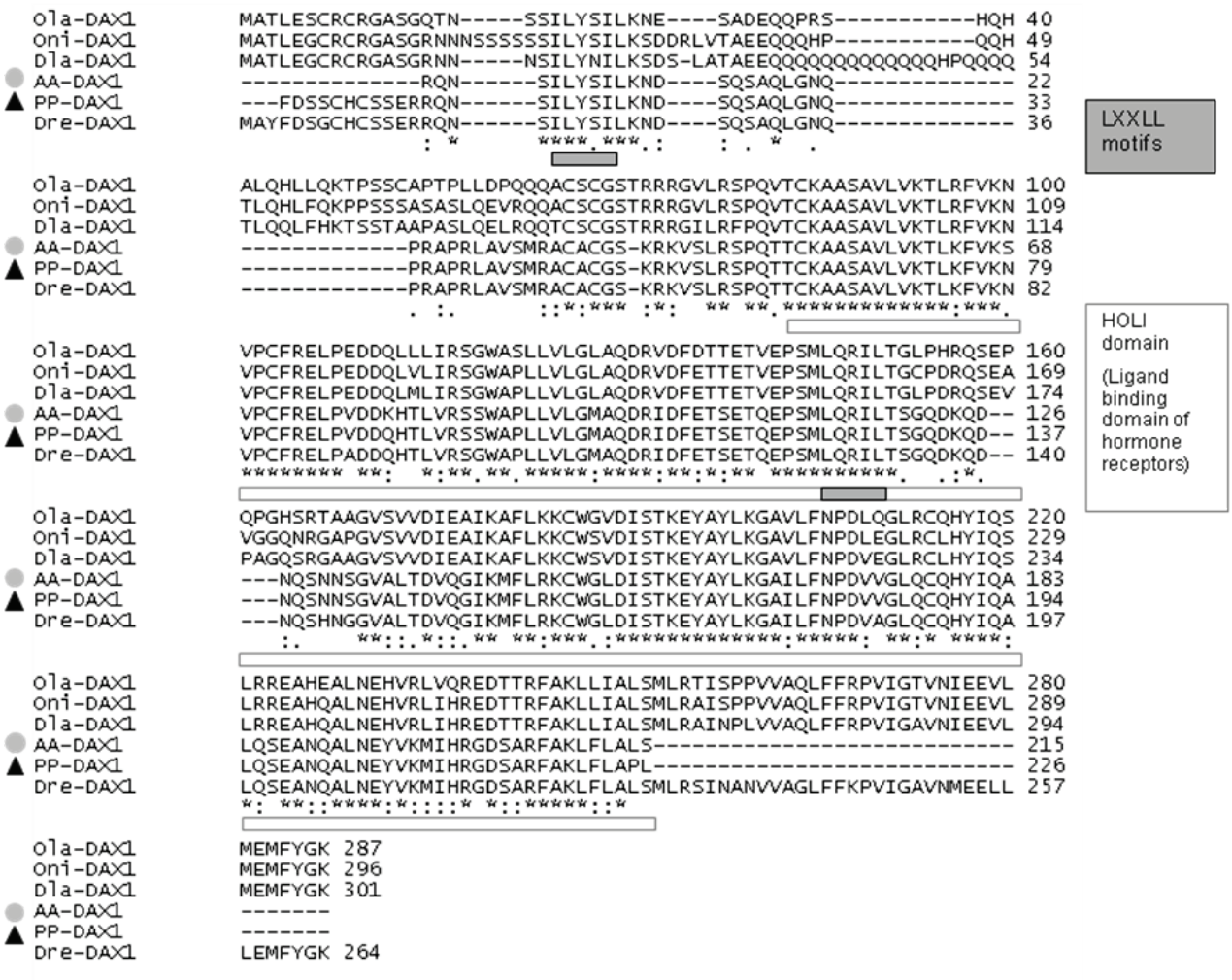


Figure 6. Protein alignment of the *Squalius* Dax1 with Dax1 orthologs of other teleosts: (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved-substitution. The LXXL motifs and the ligand binding domain that characterize Dax1 are highlighted in grey.

▲ *S. pyrenaicus* (PP) ; ● *S. alburnoides* (AA)

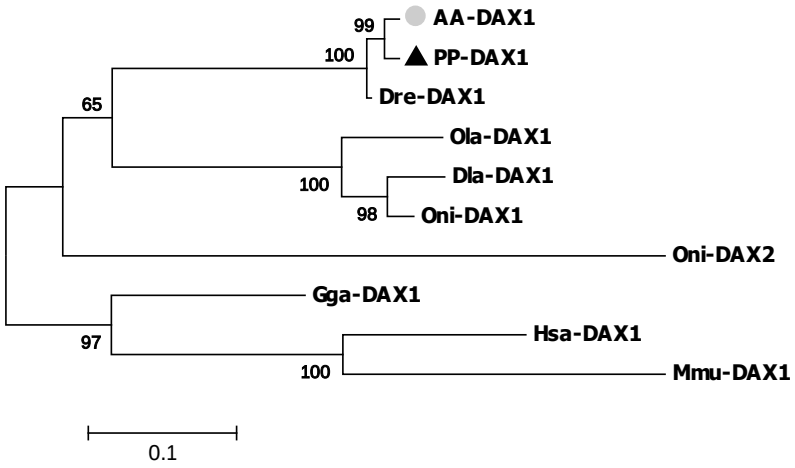


Figure 7. Unrooted Neighbour-Joining tree obtained among the DAX1 family (Dre-DAX1- *Danio rerio*, Dla-DAX1- *Dicentrarchus labrax*, Ola-DAX1- *Oryzias latipes*, Ony-DAX1- *Oreochromis niloticus,* Hsa-DAX1- *Homo sapiens*, Mmu-DAX1- *Mus musculus* and Gga-DAX1- *Gallus gallus*) and the Dax2 protein of *Oreochromis niloticus* (Oni-DAX2). Bootstrap values are shown above the branches. ▲ *S. pyrenaicus* (PP) ; ● *S. alburnoides* (AA)

3.1.4. *figla*

Fragments of approximately 300bp corresponding to a putative *figla* product were obtained using *S. alburnoides* (AA) and *S.pyrenaicus* (PP) as templates. Successful amplification was obtained both with male and female gonad samples and no sequence differences were found according to sex, which constituted a surprising finding, as *figla* was expected to be an exclusive ovarian marker and the presence of a specific transcript of this gene was only expected in female gonads. The alignment of the predicted amino acid sequences (Fig. 8) revealed a high identity with Figla proteins of several teleosts. A Helix-loop-helix domain was identified (Fig. 8), which is a diagnostic feature of the Figla transcription factor, enabling the conversion of monomers to trans-activating dimmers. The phylogenetic analysis (Fig. 9) revealed a preferential grouping with zebrafish Figla, and a closer evolutionary relatedness to other teleosts' proteins, thus confirming our isolate as a partial isolate of the *Squalius* Figla ortholog.

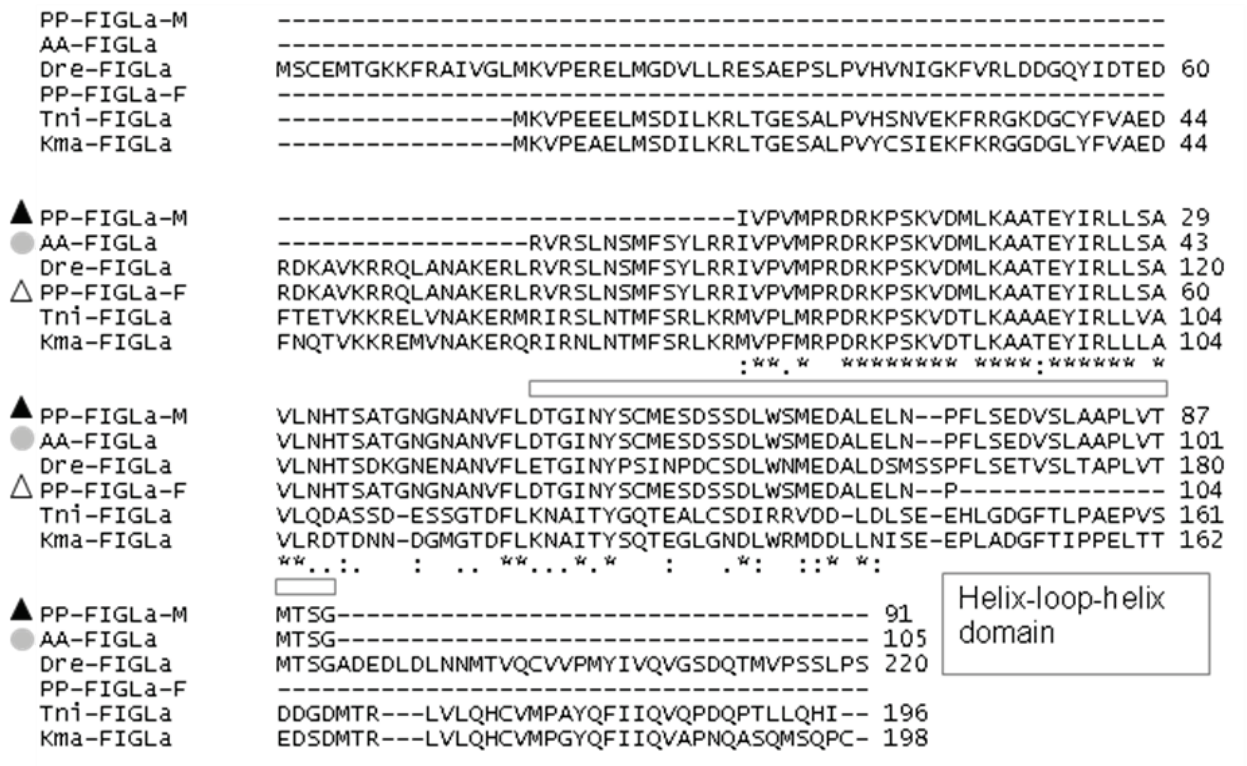


Figure 8. Protein alignment of the *Squalius* Figla with Figla orthologs of other teleosts: (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved-substitution. The helix-loop-helix domain, characteristic of the Figla proteins is highlighted in grey.

Amino acid sequences obtained from ▲ *S. pyrenaicus* males (PP-M) ; △ females (PP-F) and ● *S. alburnoides* (AA).

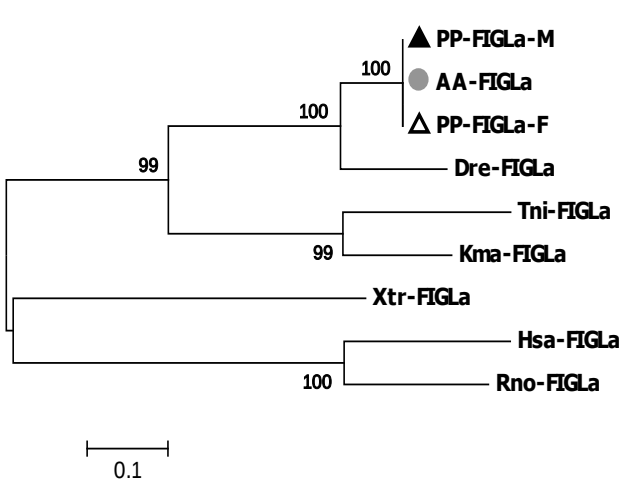


Figure 9. Unrooted Neighbour-Joining tree obtained among the FIGLA family of proteins: Dre-FIGLA- *Danio rerio*, Tni-FIGLA- *Tetraodon nigroviridis*, Kma-FIGLA- *Kryptolebias marmoratus*, Hsa- FIGLA- *Homo sapiens*, Rno-FIGLA- *Rattus norvegicus*, and Xtr FIGLA- *Xenopus (Silurana) tropicalis*. Bootstrap values are shown above the branches.

▲ *S. pyrenaicus* males (PP-M); △ females (PP-F) and ● *S. alburnoides* (AA)

3.2. Expression in the adult gonads of *S. pyrenaicus* and *S. alburnoides*

3.2.1. *dmrt1*

Expression of *dmrt1* was observed in specific cellular locations in males of the bisexual ancestor *S. pyrenaicus* (PP) and both in nuclear non-hybrid males (AA) and diploid hybrids (PA) of *S. alburnoides* (Fig. 10a, b and c). The location of *dmrt1* signal is observed in positions expected for Sertoli cells, as predicted from morphological comparison and gonad structural analysis described in Pala *et al.*, 2008a. Since the signal is more diffuse in AA gonads, we cannot exclude that *dmrt1* may also be expressed in some peripheral spermatogonia (Fig. 10b). Otherwise, no differences were observed between hybrids and parental species in terms of the location of positive sites of *dmrt1* expression, although a comparatively stronger signal was usually obtained in PP and AA gonads (Fig. 10a and b) compared to PA male hybrids (Fig 10c). *Dmrt1* expression was observed in adult ovaries of diploid (PA) and triploid (PAA) hybrid females of *S. alburnoides*. Expression was confined to the more developed cortical alveolar and yolk vesicle oocytes and was apparently absent from earlier stage oocytes (Fig 10d and e). The *dmrt1* sense probe gave no signal (Fig. 10f).

3.2.2. *wt1*

Wt1 expression in the adult testes of *S. pyrenaicus* (PP) and *S. alburnoides* (AA and PA) was observed in an area where Sertoli cells are located. No clear differences between parental and hybrid samples were observed, but again the gonads of the AA genotype appear to have a more extensive *wt1* expression, possibly in peripheral spermatogonia or interstitial cells (Fig. 10g, h and i). Very low levels of *wt1* expression were observed in adult ovaries of hybrid females of *S. alburnoides*, in oocytes at different stages of development (Fig 10j and k).

3.2.3. *dax1*

Dax1 expression was observed in adult gonads of both sexes in the *S. alburnoides* complex. Expression in males was observed exclusively outside germ cells in a position compatible with expression in Sertoli cells (Fig. 10l). In females of *S. alburnoides*, expression was observed in the adult gonad, in follicle cells surrounding perinuclear stage oocytes (Fig. 10m and n), but was absent from more mature ones, such as yolk vesicle oocytes.

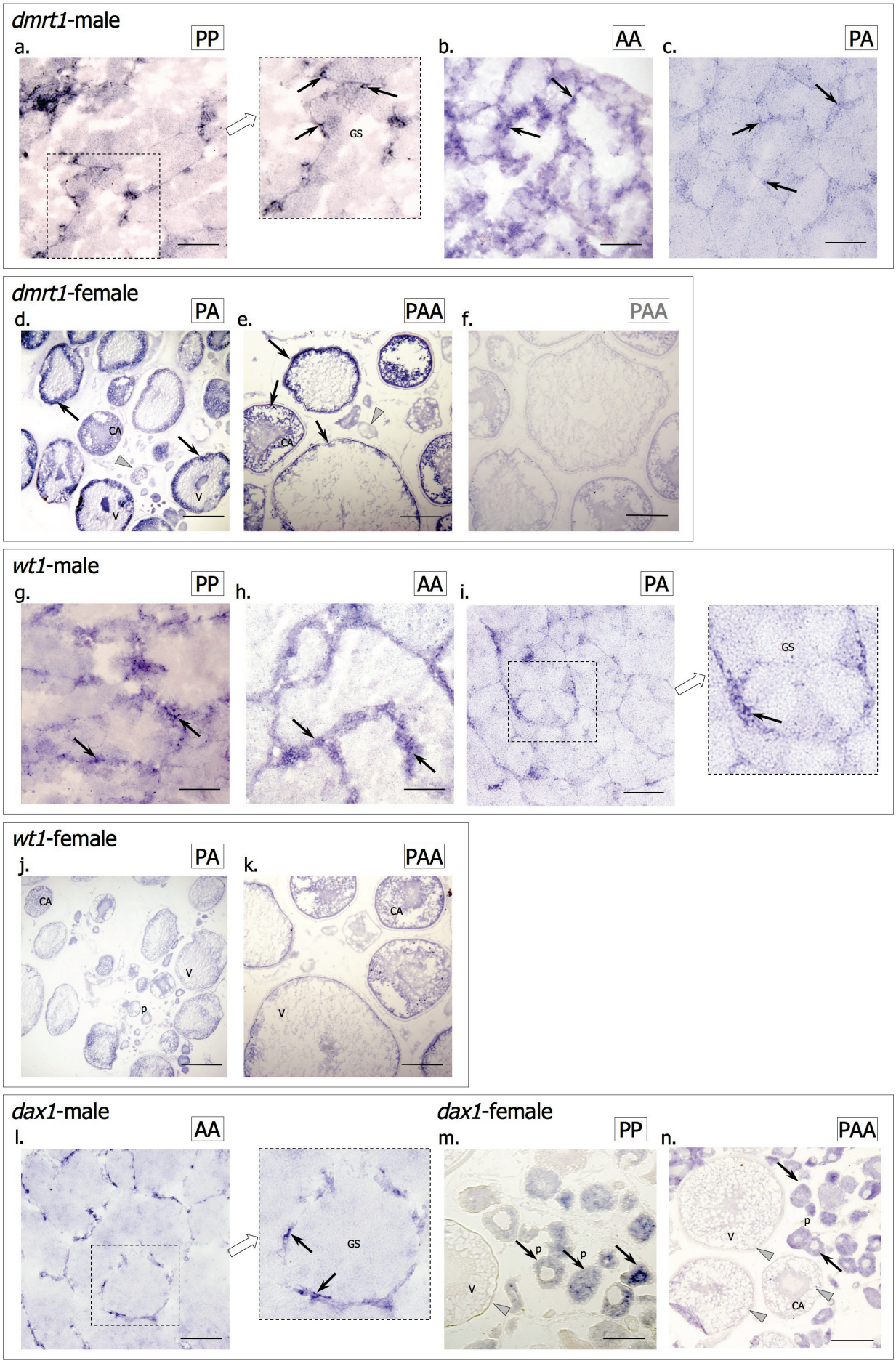

dmrt1-male
a.
PP
GS
b.
AA
c.
PA
dmrt1-female
d.
PA
CA
V
V
e.
PAA
CA
f.
PAA
wt1-male
g.
PP
h.
AA
i.
PA
GS
wt1-female
j.
PA
CA
V
P
k.
PAA
CA
V
dax1-male
l.
AA
GS
dax1-female
m.
PP
P
P
V
n.
PAA
V
P
CA

Figure 10. Expression patterns of candidate genes in the adult gonads *S. pyrenaicus* and *S. alburnoides. Dmrt1*: (a) *S. pyrenaicus* (PP genotype); (b) *S. alburnoides* (AA genotype) and (c) *S. alburnoides* (PA genotype) testis (positive signals indicated by black arrows); (d) *S. alburnoides* (PA genotype) and *S. alburnoides* (PAA genotype) ovary with (e) antisense and (f) sense probes (positive signals indicated by black arrows; cells in different maturation stages, not expressing the transcript are indicated by grey arrowheads). *Wt1*: (g) *S. pyrenaicus* (PP genotype); (h) *S. alburnoides* (AA genotype) and (i) *S. alburnoides* (PA genotype) testis (positive signals indicated by black arrows); (j) *S. alburnoides* (PA genotype) and (k) *S. alburnoides* (PAA genotype) ovary. *Dax1*: (l) *S. alburnoides* (AA genotype), (m) *S. pyrenaicus* (PP genotype) and (n) *S. alburnoides* (PAA genotype) ovary (positive signals indicated by black arrows; cells in different maturation stages, not expressing the transcript are indicated by grey arrowheads). Germ cells (GS), early perinuclear (P), cortical alveolar (CA) and vitellogenic (V) oocytes. Scale bar=100µm (a, b, e, f, g, h, k, l, m); scale bar=200µm (c, d, i, j, n).

3.2.4. *figla*

No sexual dimorphic expression of *figla* was observed in the adult *S. alburnoides*. Amplification of a *figla* specific fragment was conducted successfully both in male and female gonad cDNA templates (Fig. 11). Expression was observed in testis in locations compatible with expression in Sertoli cells (Fig. 12a). In ovaries, and as expected, *figla* was highly expressed in late primary oocytes, both in bisexual parental species *S. pyrenaicus* and in the *S. alburnoides* hybrids (Fig. 12b and c). Sections incubated with the sense probe showed no labelling (Fig. 12d).

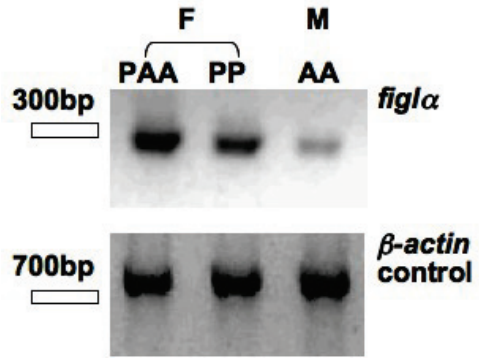


Figure 11. RT-PCR analysis of *figla* in gonad tissue of adult *S. pyrenaicus* (PP) females and *S. alburnoides* females (PAA) and males (AA). Actin controls for the same samples are shown in the lower row.

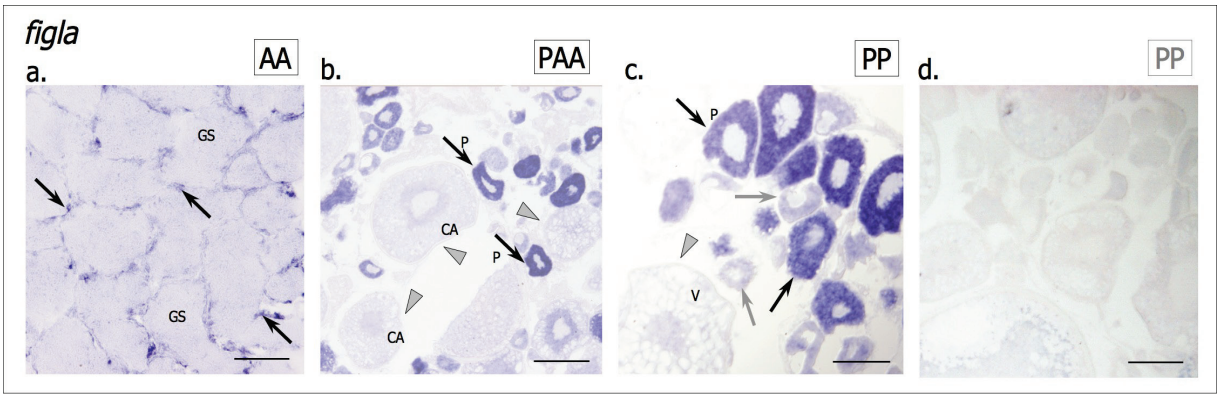


Figure 12. Expression patterns of *figla* in the adult gonads *S. pyrenaicus* and *S. alburnoides:* (a) *S. alburnoides* (AA genotype) testis; (b) *S. alburnoides* (PAA genotype) ovary and *S. pyrenaicus* (PP genotype) ovary with (c) antisense and (d) sense probes (positive signals in primary oocytes indicated by black arrows; lower expression in early stage oocytes indicated by grey arrows; cells in different maturation stages, not expressing the transcript are indicated by grey arrowheads). Germ cells (GS), early perinuclear (P), cortical alveolar (CA) and vitellogenic (V) oocytes. Scale bar=100µm (a, c, d); scale bar=200µm (b).

3.3. Expression during male development

An early role in male gonad development was confirmed for *dmrt1* in the hybrids of *S. alburnoides*. The presence of a *dmrt1* specific transcript was detected by RT-PCR at early stages in *S. alburnoides* AA progeny and its expression on the presumptive region of gonad development was observed by in situ hybridisation, as early as 6 days after hatching (dah) (Fig. 13).

A faint signal was detected for *wt1* by RT-PCR at early developmental stages. Two distinct bands were obtained from amplification with *wt1* primers using embryos as templates, as opposed to the unique band obtained from adult gonads. Expression in male embryos of AA genotype was observed at 2, 4, 6 and 14dah (Fig. 14).

Dax1 expression was not observed in early *S. alburnoides* male development at any of the stages analysed (Fig. 15).

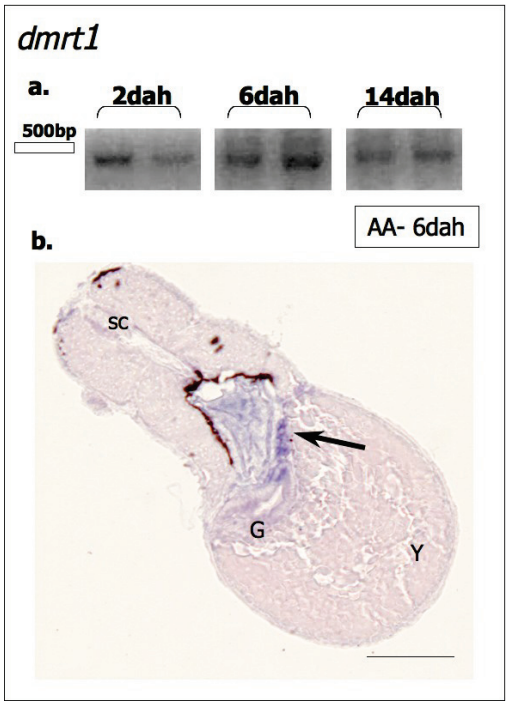


Figure 13. Expression of *dmrt1* during *S. alburnoides* development: (a) RT-PCR analysis of *dmrt1* in embryos at 2, 6 and 12 days after hatching (dah); (b) Location of *dmrt1* expression in an embryo section at 6dah in the presumptive location of the developing gonad (arrow). Spinal cord (SC); gut (G); yolk sac (Y). Scale bar=100µm.

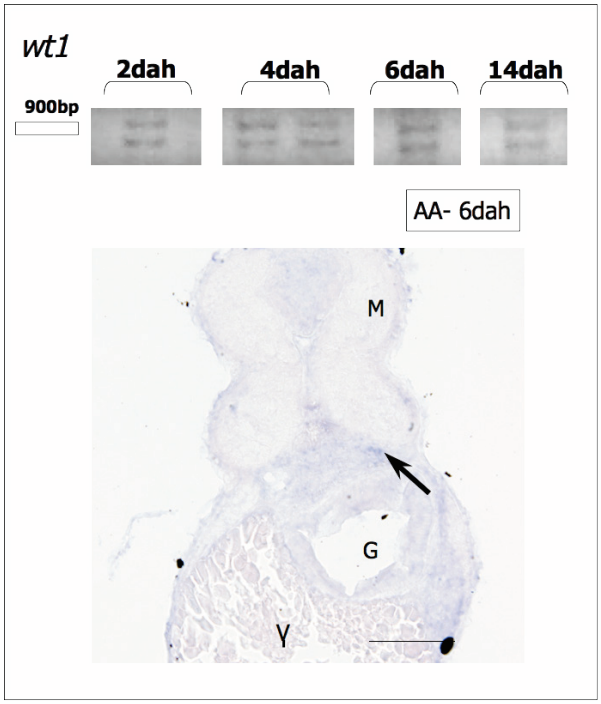


Figure 14. Expression of *wt1* during *S. alburnoides* development: (a) RT-PCR analysis of *wt1* in embryos at 2, 4, 6 and 14 days after hatching (dah); (b) Location of *wt1* expression in embryo sections at 6dah and 12dah in the presumptive location of the developing gonad (arrow). Mesoderm (M); gut (G); yolk sac (Y). Scale bar=100µm.

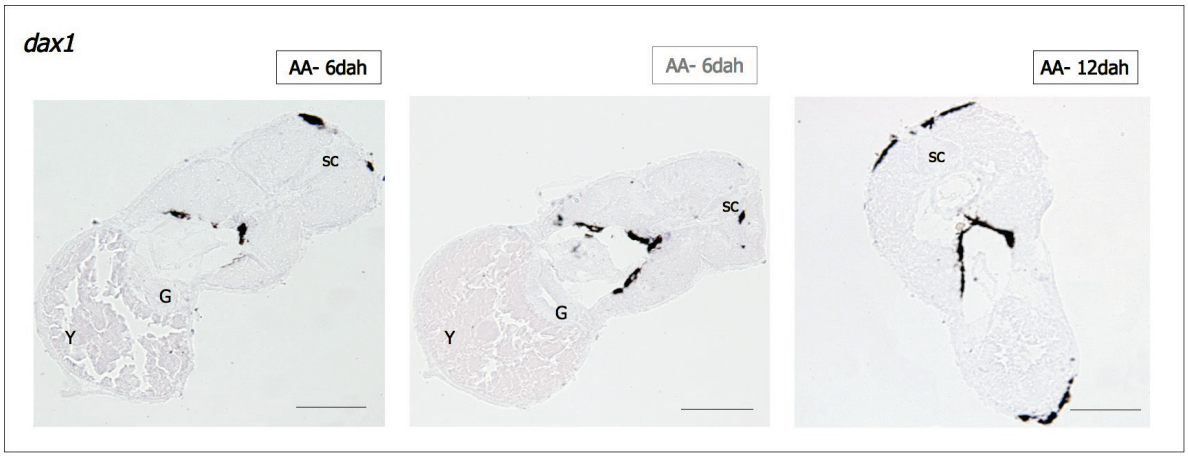


Figure 15. Absense of *dax1* positive signals in embryo sections of *S. alburnoides* (AA genotype) at 6 (sense probe in grey) and 12 days after hatching (dah). Spinal cord (SC); gut (G); yolk sac (Y). Scale bar=100µm.

Discussion

In the present study, we report the isolation of four genes known to be involved in the sex determination/ gonad differentiation pathways of mammals, predicted to play similar roles in fish, but so far not analysed in the *S. alburnoides* hybrid complex. For *dmrt1*, *wt1* and *dax1* we focused on analyzing all the characteristics that would allow us to assess their potential as possible players in male sex determination. Conversely, for *figla* the ultimate purpose was to use it as an early marker for female development. Thus, the first approach was to create a "positive" control using gonads of known sex (adults) as templates, to assess whether the reported ovary-specific expression of *figla* was also observed in *Squalius*.

In the case of *dmrt1*, our isolate presented important features for the assessment of its identity as the *Squalius* ortholog but also relevant when aiming at the production of *dmrt1* specific probes: it included not only the conserved DM domain that characterizes the family of proteins, but also a more diverse and gene specific region that unambiguously allows its distinction from other DM genes. We have reported a size polymorphism between *S. alburnoides* (A genome) and the bisexual ancestor *S. pyrenaicu*s (P genome) Dmrt1 proteins. In the presence of a genome specific difference and taking the genotype-phenotype correlation into account, an appealing hypothesis would be that homozygosity (and thus higher dosage) of the "deletion-containing" version in the AA form of *S. alburnoides* could correlate with the all-male phenotype. This would be in agreement to what has been proposed in birds, in which higher dosage of Dmrt1 (located on the Z chromosome) would be an essential factor in male sex determination (Raymond *et al.* 1999a; Smith *et al.* 1999a; Shan *et al.* 2000). However, support to the hypothesis that this putatively distinct "A-specific" form of Dmrt1, when present in double dosage, could directly contribute to the "obligatory" establishment of the male determination pathway, could only be attained if we could propose a functional or structural difference between the P and A genome versions of Dmrt1. Both the deletion and the point mutation observed in the A genome do not affect the predicted secondary structure of the protein, are outside the conserved DNA binding domain, and even though located near a putative splicing site (Guo *et al.*, 2005), apparently do not affect intron removal. We could not directly assess the effect of the polymorphisms on protein functionality, but it is also likely that there are also no relevant consequences at that level. In humans, it has been shown that mutations in the 9p DM genes (*DMRT1* and *DMRT2*) are not sufficient to cause sex reversal effects, which are usually correlated to larger regional deletions (Raymond *et al.*, 1999b). Thus, and considering the possibilities we could explore so far, we could not determine a relevant difference between "A

genome" and "P genome" *dmrt1* that could support a hypothesis placing *dmrt1* as a key regulator of male sex determination in this hybrid complex.

Regarding gene expression, *dmrt1* transcripts were found both in male and female adult gonads. The absence of a sexual dimorphic expression pattern is apparently in contrast to what has been reported for a number of fish species. In the Nile tilapia *O. niloticus*, *dmrt1* expression occurs exclusively in male gonads (Guan *et al.*, 2000) and the gene is specifically expressed in male fry very early during development, even before any signs of morphological sex differentiation (Kobayashi *et al.*, 2008). In the rainbow trout *O. mykiss*, *dmrt1* plays a role in testis development, but not in the ovary (Marchand *et al.*, 2000). In the more closely related cyprinid zebrafish though, expression patterns seem to be in agreement to what we report for *S. alburnoides* and *S. pyrenaicus*: *dmrt1* is expressed in the developing gonads of both sexes (Guo *et al.*, 2005), suggesting that it might also have an important role in ovary differentiation (as opposed to the classical view of sex dimorphic *dmrt1* expression and an exclusive association to testis development). In fact, evidence from various species suggest that timing and levels of *dmrt1* expression could be associated to sex-specific differentiation fates, with high expression levels correlating to testis development and low *dmrt1* levels being compatible with ovary differentiation (Raymond *et al.*, 2000; Nanda *et al.*, 2000; Torres Maldonado *et al.*, 2002; Shibata *et al.*, 2002).

However, differences emerge when directly comparing the cellular location of *dmrt1* expression in the adult gonads of zebrafish and of *S. alburnoides*. In males of zebrafish, *dmrt1* transcripts were identified in the developing germ cells, while in the *Squalius* samples expression is apparently restricted to Sertoli cell positions. The Sertoli cell lineage specificity and the absence of expression in the germ cell lineage are in accordance to what has been reported for the Nile tilapia (Guan *et al.*, 2000), Fugu (Yamaguchi *et al.*, 2006) and *dmrt1a* and *dmrt1bY* in the Medaka (Kobayashi *et al.*, 2004). Conversely, *dmrt1* participation in both Sertoli and germ cell lineage differentiation has been reported in the platyfish *Xiphophorus maculatus,* thus the observed differences could be a consequence of different roles in sexual development or differential gene evolution in fish sublineages (Veith *et al.*, 2006). The same could apply for gene expression patterns in females. In zebrafish, *dmrt1* gene expression is high in developing perinuclear oocytes and fainter in more mature stages (Guo *et al.*, 2005), while in *S. alburnoides* females it is mainly observed in the more developed yolk vesicle oocytes. In addition to the possibility of functional diversity within the fish group, differences in the maturation state of the ovary of each of the compared species (e.g. the low representation of perinuclear oocytes in the *Squalius* samples) could also account for the observed disparities.

Proceeding with the search for potential candidates, we have isolated a *wt1* ortholog in *S. pyrenaicus*. The sequence and phylogenetic relatedness features place it in close proximity to the *wt1a* gene of zebrafish. The latter species possesses a second *wt1* gene, *wt1b* and both paralogous show overlapping but distinctive spatiotemporal expression in kidney and gonads (Bollig *et al.*, 2006), playing an important role in pronephric differentiation. In the present work we were only able to identify one *wt1* isolate in *Squalius*, despite having tested primer combinations based on the two zebrafish genes (*wt1a* and *wt1b*), but we cannot exclude the possibility of presence of a second wt1 gene in the Iberian species. Our isolate included a conserved *wt1* specific region and two of the four zinc fingers that are necessary for its functions as a bona fide transcription factor and regulator of RNA processing (reviewed in Englert, 1998), thus providing good specificity as template for hybridisation probe design. *Wt1* probes generated from *S. pyrenaicus* were fully functional in *S. alburnoides*, implying sequence identity between genomes.

Wt1 expression was observed in the adult gonads of male and female *S. pyrenaicus* and *S. alburnoides*. In males, *wt1* is expressed in the Sertoli cell lineage and apparently absent form germ cells, which is in agreement to what has been reported in other vertebrates. Although *wt1* orthologs in fish have been studied in relation to pronephric development, a role of *wt1* in the determination, differentiation and identity maintenance of the Sertoli cell lineage was assessed in mice (Gao *et al.*, 2006) through the identification of gene targets activated by *wt1* that are known to be expressed in Sertoli cells of developing and adult gonads. Thus, it is possible that *wt1* is playing a similar function in the male gonads of the *Squalius* cyprinids. The apparently lower expression in female gonads could also be related to the role of *wt1* in follicle development. In different mammals and birds (Chun *et al.*, 1999), higher expression in early stage oocytes and a progressively lowering in *wt1* expression during follicle maturation implied an additional role of this gene: the repression of genes involved in follicle differentiation, and the maintenance of a number of oocytes that can later on be recruited, in early developmental stages. It is not clear whether or not this could be the role that *wt1* is playing in *S. alburnoides* ovary but the fact that most oocytes present in the analysed samples represent later stages of maturation and show low *wt1* expression, would be in accordance to that possibility. A more detailed analysis, using ovary samples in different maturation stages would be necessary to fully understand the role of *wt1* transcripts in the female gonads of this complex and provide further insight into the process of female gonad development in fish.

Regarding *dax1*, the isolates obtained from *S. alburnoides* and *S. pyrenaicus* were shown to be conserved within the cyprinid group, with high similarity to the zebrafish protein. Specific structural domains, characteristic of Dax1 and essential for its participation in signal transduction cascades

were present in the *Squalius* isolates, thus implying similar functionality and potential for interaction with similar partners of the *Squalius* Dax1 proteins.

In the adult *Squalius* gonads, *dax1* shows no sexual dimorphic expression pattern, which is in accordance to what has been described in a number of other teleost species. No sex specific expression differences were found between female and male adult gonads in the nile tilapia *O. niloticus* (Wang *et al.*, 2002) and throughout the gonadal sex differentiation period in the European sea bass *D. labrax* (Martins *et al.*, 2007). Conversely, in the medaka, *dax1* mRNA was absent from the adult testis, but expression was detected in postvitellogenic follicles of the adult ovary (Nakamoto *et al.*, 2007). In *S. alburnoides* and *S. pyrenaicus* expression of *dax1* in the adult gonad is apparently restricted to Sertoli cells and absent from the germ cell lineage in males. In females, *dax1* transcripts are mainly observed in granulosa cells surrounding perinuclear oocytes. The expression patterns of this gene in the *Squalius* gonads could well correlate with the functions of modulation of the activity of other factors and participation in the maintenance of integrity of testis and ovary structure and organization, attributed to *dax1* (Clipsham & McCabe, 2003). In fact, it has been proposed that the effects observed in Dax1 mutants in mouse, affecting normal granulosa cell organization around oocytes in females, could be functionally related to the defects in Sertoli cell support of germ cells observed in males (Yu *et al.*, 1998). Taking the expression patterns of *dax1* in the same cell types into account, it is possible that in the adult gonads of both sexes, *dax1* could also be contributing to the regulation of supporting cell lineages and indirectly modulating gametogenesis in the *Squalius* gonads.

For all candidates, we have performed an initial assessment of whether, besides their contribution to regulation in the adult gonads, they might also have a direct role in the earlier process of sex determination. The evaluation of a putative role and consequent interest for further analysis would be based on whether they were expressed in embryos and their contribution was early during development, and if we could observe expression in the developing gonadal structures. As no sex marker is available for the *Squalius* species, we selected the putative all-male progeny of AA genotype to perform the test, avoiding the ambiguities that could result from the use of a diversity of individuals of unknown sex. A good certainty of the embryos' sex would allow us to determine whether the results obtained would be in agreement to the role described for the three genes in other species, during sexual development. We observed that, in agreement to other teleosts (Kobayashi *et al.*, 2008; Marchand *et al.*, 2000, Bollig *et al.*, 2006), both *dmrt1* and *wt1* are expressed in *S. alburnoides* male progeny, at early stages and in the developing gonad, thus implying a contribution to its differentiation. However, the real participation of *wt1* and *dmrt1* as sex determination genes can only be assessed when male and female embryos can be analysed in

parallel. The same is true regarding *dax1* expression. No *dax1* transcripts were observed in the early developmental stages of *S. alburnoides* males. It is possible that levels of *dax1* expression might vary during development, as shown in mouse (Ikeda *et al.*, 2001) and that the stages analysed in *S. alburnoides* could correspond to "windows" of *dax1* downregulation. However, the absence of *dax1* expression at early developmental stages apparently implies a lack of involvement in male sex determination, nevertheless maintaining the possibility of a later enrolment in differentiation processes.

Regarding *figla*, and its potential use as an early ovary marker in *Squalius*, the expression results obtained in adult gonads imply a completely opposite scenario to the one initially predicted. Female restricted *Figla* expression has been described for a number of species, namely in mammals and fish (Soyal *et al.*, 2000), (Jørgensen *et al.*, 2008), so it was unexpected to isolate a *figla* transcript from *S. pyrenaicus* and *S. alburnoides* adult testis. No differences were found between female and male *figla* isolates in terms of sequence, domain prediction and putative functionality.

In fish like the gilthead seabream *Sparus aurata*, sex change occurs at a certain stage of the life cycle, from a heterosexual gonad, with gene expression profiles changing accordingly (Liarte *et al.*, 2007). Thus, an initial hypothesis would be that gonad maturation in *S. alburnoides* could also involve a similar process and that Figla transcripts detected by RT-PCR could originate from a vestigial female structure. Data from *in situ* hybridisation proved this hypothesis to be remote. *Figla* transcripts were detected in Sertoli cell locations in males with normal testis structural organization and with no morphological hints of the presence of any female specific gonadal structure. In females, on the other hand, *figla* expression apparently fits the expectations. Expression was restricted to cells surrounding early stage oocytes, which is in close agreement with what has been reported in the medaka (Kanamori, 2000) and is consistent with the role of *figla* in follicle development. A more reasonable possibility to account for non-dimorphic expression would be the presence of *figla* transcripts both in *Squalius* males and females, with differential expression levels of *figla* during development participating in the triggering of male or female differentiation. In fact, a peak in *figla* expression has been correlated to ovary differentiation in zebrafish (Jørgensen *et al.*, 2008). An alternative hypothesis could be that *figla* female specific expression would only occur during development and that in the adult it would be expressed in cells supporting the germ cell lineage. These possibilities can only be tested more systematically when a straightforward identification of male and female *Squalius* embryos becomes possible and the expression patterns can be followed in parallel in both sexes. Additionally, it would be of interest to follow expression of *figla* and of the other sex determination candidates quantitatively, at different developmental stages and in levels of gonad maturation in the adult, as timing and expression levels of genes in sex determination

cascades have been shown to be critical to promote the switch between male and female determination pathways (e.g. Raymond *et al.*, 2000; Nanda *et al.*, 2000; Torres Maldonado *et al.*, 2002; Shibata *et al.*, 2002; Jørgensen *et al.*, 2008).

A more comprehensive analysis of the putative roles and interactions of the isolated candidates is currently hampered by the impossibility of unambiguous sexing of embryos. While trying to overcome this difficulty, our aim was to create the framework that could subsequently be extended to further analysis: candidates were characterized and compared, gene specific probes were synthesized and tested, cellular and organ specific locations were assessed. In our point of view, by establishing a working basis of different genes, the present work constitutes the necessary initial step towards the study of the genetic basis of such a complex process in such an unorthodox hybrid system.

References:

Alves, M.J., Coelho, M.M. & Collares-Pereira, M.J., 1998. Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. Mol. Biol. Evol.15, 1233-1242.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111, 375–385.

Armstrong, J.F., Pritchard-Jones, K., Bickmore, W.A., Hastie, N.D., Bard, J.B., 1993. The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. Mech Dev.40, 85-97.

Bajanca, F., Luz, M., Duxson, M.J., Thorsteinsdóttir, S., 2004. Integrins in the mouse myotome: developmental changes and differences between the epaxial and hypaxial lineage. Dev. Dyn. 231, 402–415.

Bardoni, B., Zanaria, E., Guioli, S., Floridia, G., Worley, K.C., Tonini, G, Ferrante, E., Chiumello, G., McCabe, E.R., Fraccaro, M., Zuffardi, O., Camerino, G., 1994. A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. Nat Genet 7, 497–501.

Baroiller, J.-F., 1999. Endocrine and environmental aspects of sex differentiation in fish. Cellular and Molecular Life Sciences (CMLS) 55, 910–931.

Bollig, F., Mehringer, R., Perner, B., Hartung, C., Schäfer, M., Schartl, M., Volff, J.N., Winkler, C., Englert, C., 2006. Identification and comparative expression analysis of a second wt1 gene in zebrafish. Dev Dyn. 235, 554-561.

Bor, Y.C.. Swartz, J., Morrison, A., Rekosh, D., Ladomery, M., Hammarskjold, M.L., 2006. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. Genes Dev 20, 1597–1608.

Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T., 2005. Protein structure prediction servers at University College London. Nucl. Acids Res. 33 (Web Server issue), W36-38.

Brunelli, J.P., Robison, B.D., Thorgaard, G.H., 2001. Ancient and recent duplications of the rainbow trout Wilms' tumor gene. Genome 44, 455-462.

Chun, S.Y., McGee, E.A., Hsu, S.Y., Minami, S., LaPolt, P.S., Yao, H.H., Bahr, J.M., Gougeon, A., Schomberg, D.W., Hsueh, A.J., 1999. Restricted expression of WT1 messenger ribonucleic acid in immature ovarian follicles: uniformity in mammalian and avian species and maintenance during reproductive senescence. Biol Reprod. 60, 365-373.

Clipsham, R., McCabe, E.R., 2003. DAX1 and its network partners: exploring complexity in development. Mol Genet Metab.80, 81-120.

Collares-Pereira, M.J., Próspero, M. I., Biléu, R.I., Rodrigues, E. M., 1998. *Leuciscus* (Pisces, Cyprinidae) karyotypes: transect of Portuguese populations. Gen. Mol. Biol. 21, 63-69.

Devlin, R.H., Nagahama, Y., 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208, 191–364.

Englert, C., 1998. WT1- more than a transcription factor? Trends Biochem Sci 23, 389-393.

Erdman, S.E., Burtis, K. C., 1993. The *Drosophila doublesex* proteins share a novel zinc finger related DNA binding domain. EMBO J. 12, 527–535.

Fedorova, S., Miyamoto, R., Harada, T., Isogai, S., Hashimoto, H., Ozato, K., Wakamatsu, Y., 2008. Renal glomerulogenesis in medaka fish, Oryzias latipes. Dev Dyn. 237, 2342-2352.

Ferguson-Smith, M., 2007. The evolution of sex chromosomes and sex determination in vertebrates and the key role of DMRT1. Sex Dev 1, 2-11.

Gao, F., Maiti, S., Alam, N., Zhang, Z., Deng, J.M., Behringer, R.R., Lécureuil, C., Guillou, F., Huff, V., 2006. The Wilms tumor gene, Wt1, is required for Sox9 expression and maintenance of tubular architecture in the developing testis. Proc Natl Acad Sci U S A. 103, 11987-11992.

Guan, G., Kobayashi, T., Nagahama, Y., 2000. Sexually dimorphic expression of two types of DM (Doublesex/Mab-3)-domain genes in a teleost fish, the Tilapia (*Oreochromis niloticus*).Biochem Biophys Res Commun. 272, 662-666.

Guo, Y., Cheng, H., Huang, X., Gao, S., Yu. H., Zhou, R., 2005. Gene structure, multiple alternative splicing, and expression in gonads of zebrafish Dmrt1. Biochemical and Biophysical Research Communications 330, 950–957.

Holter, E., Kotaja, N., Makela, S., Strauss, L., Kietz, S., Janne, O.A., Gustafsson, J.A., Palvimo, J.J., Treuter, E., 2002. Inhibition of androgen receptor (AR) function by the reproductive orphan nuclear receptor DAX- 1. Mol Endocrinol 16, 515-528.

Ikeda, Y., Takeda, Y., Shikayama, T., Mukai, T., Hisano, S., Morohashi, K. I., 2001. Comparative localization of Dax-1 and Ad4BP/SF-1 during development of the hypothalamic-pituitary-gonadal axis suggests their closely related and distinct functions. *Dev. Dyn.* 220, 363- 376.

Iyer, A.K., McCabe, E.R., 2004. Molecular mechanisms of DAX1 action. Mol Genet Metab 83, 60-73.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195-202.

Jørgensen, A., Morthorst, J.E., Andersen, O., Rasmussen, L.J., Bjerregaard, P., 2008. Expression profiles for six zebrafish genes during gonadal sex differentiation. Reprod Biol Endocrinol. 6, 25.

Kanamori, A., 2000. Systematic identification of genes expressed during early oogenesis in medaka. Mol Reprod Dev. 55, 31-36.

Kent, J., Coriat, A.M., Sharpe, P.T., Hastie, N.D., van Heyningen, V., 1995. The evolution of WT1 sequence and expression pattern in the vertebrates. Oncogene 11, 1781-1792.

Kim, J., Prawitt, D., Bardeesy, N., Torban, E., Vicaner, C., Goodyer, P., Zabel, B., Pelletier, J., 1999.The Wilms' tumor suppressor gene (wt1) product regulates Dax-1 gene expression during gonadal differentiation. Mol Cell Biol 19, 2289–2299.

Klüver, N., Herpin, A., Braasch, I., Drieβle, J., Schartl, M., 2008. Regulatory back-up circuit of medaka Wt1 co-orthologs ensures PGC maintenance. Dev Biol. [Epub ahead of print]

Kobayashi, T., Matsuda, M., Kajiura-Kobayashi, H., Suzuki, A., Saito, N., Nakamoto, M., Shibata, N., Nagahama, Y., 2004. Two DM domain genes, *DMY* and *DMRT1*, involved in testicular differentiation and development in the medaka, *Oryzias latipes*. Dev Dyn 231, 518–526.

Kobayashi, T., Kajiura-Kobayashi, H., Guan, G., Nagahama, Y., 2008. Sexual dimorphic expression of DMRT1 and Sox9a during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*). Dev Dyn. 237, 297-306.

Lambert, C., Leonard, N., De Bolle, X., Depiereux, E., 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18, 1250-1256.

Letunic, I., Doerks, T., Bork, P., 2009. SMART 6: recent updates and new developments.Nucleic Acids Res. 37(Database issue):D229-232. [Epub 2008]

Liang, L., Soyal, S.M., Dean, J., 1997. FIGalpha, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes. Development 124, 4939-4947.

Liarte, S., Chaves-Pozo, E., García-Alcazar, A., Mulero, V., Meseguer, J., García-Ayala, A., 2007. Testicular involution prior to sex change in gilthead seabream is characterized by a decrease in DMRT1 gene expression and by massive leukocyte infiltration. Reprod Biol Endocrinol. 5, 20.

Maldonado, L.C.T., Piedra, A.L., Mendoza, N.M., Valencia, A.M., Martinez, A.M., Larios, H.M., 2002. Expression Profiles of Dax1, Dmrt1, and Sox9 During Temperature Sex Determination in Gonads of the Sea Turtle Lepidochelys Olivacea. General and Comparative Endocrinology 129, 20-26.

Manolakou, P., Lavranos, G., Angelopoulou. R., 2006. Molecular patterns of sex determination in the animal kingdom: a comparative study of the biology of reproduction. Reprod Biol Endocrinol. 4, 59.

Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J., Bernot, A., Laudet, V., Guiguen, Y., 2000. DMRT1 expression during gonadal differentiation and spermatogenesis in therainbow trout, *Oncorhynchus mykiss*. Biochim Biophys Acta 1493, 180-187.

Martins, R.S., Deloffre, L.A., Mylonas, C.C., Power, D.M., Canário, A.V., 2007. Developmental expression of DAX1 in the European sea bass, Dicentrarchus labrax: lack of evidence for sexual dimorphism during sex differentiation. Reprod Biol Endocrinol. 5, 19.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., Sakaizumi, M., 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417, 559– 563.

Nachtigal, M.W., Hirokawa, Y., Enyeart-VanHouten, D.L., Flanagan, J.N., Hammer, G.D., Ingraham, H.A., 1998. Wilms' tumor 1 and DAX-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. Cell 93, 445-454.

Nakamoto, M., Wang, D.S., Suzuki, A., Matsuda, M., Nagahama, Y., Shibata, N., 2007. Dax1 suppresses P450arom expression in medaka ovarian follicles. Mol Reprod Dev. 74, 1239-1246.

Nakatsuru, Y., Minami, K., Yoshikawa, A., Zhu, J., Oda, H., Masahito, P., Okamoto, N., Nakamura, Y., Ishikawa, T., 2000. Eel WT1 sequence and expression in spontaneous nephroblastomas in Japanese eel. Gene 245, 245-251.

Nanda, I,, Shan, Z., Schartl, M., Burt, D.W., Koehler, M., Nothwang, H., Grützner, F., Paton, I.R.,Windsor, D., Dunn, I., Engel, W., Staeheli, P., Mizuno, S., Haaf, T., Schmid, M., 1999. 300 million years of conserved synteny between chicken Z and human chromosome 9. Nat Genet. 21, 258-259.

Nanda, I., Zend-Ajusch, E., Shan, Z., Grutzner, F., Schartl, M., Burt, D.W., Koehler, M., Fowler, V.M., Goodwin, G., Schneider, W.J., Mizuno, S., Dechant, G., Haaf, T., Schmid, M., 2000. Conserved Synteny Between the Chicken Z Sex Chromosome and Human Chromosome 9 Includes the Male Regulatory Gene Dmrt1: a Comparative (Re)View on Avian Sex Determination. Cytogenetics and CellGenetics 89, 67-78.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., Schmid, M., Schartl, M., 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. Proceedings of the National Academy of Science USA 99, 11778 – 11783.

Niksic, M., Slight, J., Sanford, J.R., Caceres, J.F., Hastie, N.D., 2004. The Wilms' tumour protein (WT1) shuttles between nucleus and cytoplasm and is present in functional polysomes. Hum Mol Genet 13, 463–471.

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401-437.

Pala, I., Klüver, N., Thorsteinsdottir, S., Schartl, M., Coelho, M.M., 2008a. Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*. Gene 410, 249-258.

Pala, I., Coelho, M.M, Schartl, M., 2008b. Dosage compensation by gene copy silencing in a triploid hybrid fish. Curr Biol. 18, 1344-1348.

Perner, B., Englert, C., Bollig, F., 2007. The Wilms tumor genes wt1a and wt1b control different steps during formation of the zebrafish pronephros. Dev Biol. 309, 87-96.

Raymond C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J., Zarkower, D., 1998. Evidence for evolutionary conservation of sex-determining genes. Nature 391, 691-695.

Raymond, C.S., Kettlewell, J.R, Hirsch, B., Bardwell, V.J., D., Zarkower, D., 1999a. Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. Dev. Biol. 215, 208–220.

Raymond, C.S., Parker, E.D., Kettlewell, J.R., Brown, L.G., Page, D.C., Kusz, K., Jaruzelska, J., Reinberg, Y., Flejter, W.L., Bardwell, V.J., Hirsch, B., Zarkower, D., 1999b. A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. Hum Mol Genet. 8, 989-996.

Raymond, C.S., Murphy, M.W., O'sullivan, M.G., Bardwell, V.J., Zarkower, D., 2000. Dmrt1, a Gene Related to Worm and Fly Sexual Regulators, Is Required for Mammalian Testis Differentiation. Genes & Development 14, 2587-2595.

Rivera, M.N., Haber, D.A., 2005. Wilms' tumour: connecting tumorigenesis and organ development in the kidney. Nature Reviews Cancer 5, 699-712.

Rodriguez-Mari, A., Yan, Y.L., Bremiller, R.A., Wilson, C., Canestro, C., Postlethwait, J.H., 2005. Characterization and expression pattern of zebrafish Anti-Mullerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. Gene Expr. Patterns 5, 655–667.

Schartl, M., 2004.A comparative view on sex determination in medaka. Mech Dev. 121, 639-645.

Semba, K., Saito-Ueno, R., Takayama, G., Kondo, M., 1996. cDNA cloning and its pronephros-specific expression of the Wilms' tumour suppressor gene, WT1, from *Xenopus laevis*. Gene 175, 165-172.

Shan, Z., Nanda, I., Wang, Y., Schmid, M., Vortkamp, A., Haaf, T., 2000. Sex-specific expression of an evolutionarily conserved male regulatory gene, DMRT1, in birds, Cytogenet. Cell Genet. 89, 252–257.

Shibata, K., Takase, M., Nakamura, M., 2002. The Dmrt1 Expression in Sex- Reversed Gonads of Amphibians. General and Comparative Endocrinology 127, 232-241.

Smith, C.A., McClive, P. J., Western, P.S., Reed, K.J., Sinclair, A. H., 1999a. Conservation of a sex-determining gene, Nature 402, 601– 602.

Smith, C.A., Smith, M.J., Sinclair, A.H., 1999b. Gene expression during gonadogenesis in the chicken embryo. Gene 234, 395-402.

Smith, C.A., Clifford, V., Western, P.S., Wilcox, S.A., Bell, K.S., Sinclair, A,H., 2000. Cloning and expression of a DAX1 homologue in the chicken embryo. Gene 241, 223-232.

Smith, C.A., Katz, M., Sinclair, A.H., 2003. DMRT1 is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. Biol Reprod. 68, 560-570.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V.C., 2006. May a hybridogenetic complex regenerate the nuclear genome of both sexes of a missing ancestor? First evidence on the occurrence of a nuclear non-hybrid Squalius alburnoides (Cyprinidae) female based on DNA sequencing. Journal of Natural History 40, 1443–1448.

Soyal, S.M., Amleh, A., Dean, J., 2000. FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. Development 127, 4645-4654.

Sugita, J., Takase, M., Nakamura, M., 2001. Expression of Dax-1 during gonadal development of the frog. Gene 280, 67-74.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24,1596-1599.

Torres Maldonado, L.C., Landa Piedra, A., Moreno Mendoza, N., Marmolejo Valencia, A., Meza Martínez, A., Merchant Larios, H., 2002. Expression profiles of Dax1, Dmrt1, and Sox9 during temperature sex determination in gonads of the sea turtle Lepidochelys olivacea. Gen Comp Endocrinol. 129, 20-26.

Trant, J.M., Gavasso, S., Ackers, J., Chung, B.C., Place, A.R., 2001.Developmental expression of cytochrome P450 aromatase genes (CYP19a and CYP19b) in zebrafish fry (Danio rerio). J Exp Zool. 290, 475-483.

Tremblay, J.J., Viger, R.S., 2001. Nuclear receptor Dax-1 represses the transcriptional cooperation between GATA-4 and SF-1 in Sertoli cells. Biol Reprod. 64, 1191-1199.

Veith, A.M., Schäfer, M., Klüver, N., Schmidt, C., Schultheis, C., Schartl, M., Winkler, C., Volff, J.N., 2006. Tissue-specific expression of dmrt genes in embryos and adults of the platyfish *Xiphophorus maculatus*. Zebrafish 3, 325-337.

Veitia, R.A., Salas-Cortés, L., Ottolenghi, C., Pailhoux, E., Cotinot, C., Fellous, M., 2001. Testis determination in mammals: more questions than answers. Mol Cell Endocrinol. 179, 3-16.

von Hofsten, J., Olsson, P.E., 2005. Zebrafish sex determination and differentiation: involvement of FTZ-F1 genes. Reprod Biol Endocrinol. 3, 63.

Wang, Z. J., Jeffs, B., Ito, M., Achermann, J.C., Yu, R.N., Hales, D.B., Jameson, J.L., 2001. Aromatase (Cyp19) expression is up-regulated by targeted disruption of Dax1. Proc Natl Acad Sci USA 98, 7988-7993.

Wang, D.S., Kobayashi, T., Senthilkumaran, B., Sakai, F., Sudhakumari, C.C., Suzuki, T., Yoshikuni, M., Matsuda, M., Morohashi, K., Nagahama, Y., 2002. Molecular cloning of DAX1 and SHP cDNAs and their expression patterns in the Nile tilapia, *Oreochromis niloticus*. Biochem Biophys Res Commun. 297, 632-640

Western, P.S., Harry, J.L., Marshall Graves J.A., Sinclair, A.H., 2000. Temperature-dependent sex determination in the American alligator: expression of SF1, WT1 and DAX1 during gonadogenesis. Gene 241, 223-232.

Wilhelm, D., Englert, C., 2002. The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1 . Genes Dev 16, 1839–1851.

Yamaguchi, A., Lee, K.H., Fujimoto, H., Kadomura, K., Yasumoto, S., Matsuyama, M., 2006. Expression of the *DMRT* gene and its roles in early gonadal development of the Japanese pufferfish *Takifugu rubripes*. Comp Biochem Physiol D 1, 59–68.

Yu, R.N., Ito, M., Saunders, T.L., Camper, S.A., Jameson, J.L., 1998. Role of Ahch in gonadal development and gametogenesis. Nat Genet. 20, 353-357.

Zhang, H., Thomsen, J.S., Johansson, L., Gustafsson, J.A., Treuter, E., 2000. DAX- 1 functions as an LXXLL-containing corepressor for activated estrogen receptors. J Biol Chem 2000, 275, 39855-39859.

Zhao, Y., Yang, Z., Phelan, J.K., Wheeler, D.A., Lin, S., McCabe, E.R., 2006. Zebrafish dax1 is required for development of the interrenal organ, the adrenal cortex equivalent. Mol Endocrinol. 20, 2630-2640.

CHAPTER 4

Dosage compensation by gene copy silencing in a triploid hybrid fish

Irene Pala, Maria Manuela Coelho, Manfred Schartl

Current Biology 19, 1344-1348 (2008)

Current Biology *18*, 1344–1348, September 9, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.07.096

Report

Dosage Compensation by Gene-Copy Silencing in a Triploid Hybrid Fish

Irene Pala,[1,2] Maria M. Coelho,[1] and Manfred Schartl[2,*]
[1]Universidade de Lisboa
Centro de Biologia Ambiental/Departamento de Biologia Animal
Faculdade de Ciências de Lisboa
1749-16 Lisbon
Portugal
[2]Physiologische Chemie I
Biozentrum
University of Würzburg
97074 Würzburg
Germany

Summary

In mammals, the increase in gene dosage, in the form of polyploidy or involving chromosomal fragments, has deleterious effects [1]. Regulation of appropriate gene product amounts has to be warranted by complex dosage-compensation mechanisms. Lower vertebrates, on the other hand, cope very well with ploidy increase [2–4], implying either effective compensation or a lack of necessity for such mechanisms. Unfortunately, nothing is known about the genetic and molecular mechanisms underlying this phenomenon. For an experimental approach, we have studied gene expression in the allotriploid form of *Squalius alburnoides*. In these organisms, different genomes are joined through hybridization; thus, sequence differences can be used to follow expression of different alleles [5, 6]. We found that a compensation mechanism exists, reducing transcript levels to the diploid state. Our data suggest a silencing of one of the three alleles. Unexpectedly, it is not a whole haplome that is inactivated. The allelic expression patterns differ between genes and between different tissues for one and the same gene. Our data provide the first evidence of a regulation mechanism involving gene-copy silencing in a triploid vertebrate.

Results and Discussion

To investigate whether a regulation mechanism is operating on gene dosage in a polyploid genome context, we used triploids of the *S. alburnoides* complex. Triploids are the most predominant and widespread form of this cyprinid fish complex, which is composed of animals of different ploidy degrees and includes fertile sexual as well as rarer asexual forms (Figure 1) [7, 8]. To determine whether gene dosage is compensated at the expression level, we first tried to estimate the ratio between amounts of the *β-actin* transcript to *β-actin* gene dosage by PCR and RT-PCR of coextracted DNA and RNA from liver of diploid and triploid individuals. Although difficult to numerically quantify because of the limitations by the material, a robust tendency was observed for the RNA:DNA ratio to be lower in triploids (Figure S2, available online). Then the amount of transcripts in muscle, eye, liver, and gonad of diploid and triploid fish was more precisely determined by quantitative real-time RT-PCR analysis. The experimental procedure involved an initial normalization step assuming a 1:1 DNA:RNA ratio in diploids and triploids, although it should be considered that the total amount of mRNA, which constitutes only a minor fraction of total RNA, could be increased in triploids compared to diploids. The analysis was performed for the housekeeping genes *β-actin*, *rpl8* (*ribosomal protein l8*), and *gapdh* (*glyceraldehyde-3-phosphatedehydrogenase*) and the gonad-specific genes *amh*, *dmrt1*, and *vasa*. No significant differences were found in expression levels between fish of different ploidy and genome constitution, for all genes analyzed, both in somatic organs and gonads. The relative expression ratios obtained from the comparison of diploid (AA, PP, and PA) to triploid (PAA) samples were always approximately 1 (Table S1). In summary, this indicated a possible dosage-compensation mechanism in triploid fish, reducing expression to the diploid level.

An obvious possibility to account for dosage compensation could be the transcriptional silencing of a whole genome in triploids. We therefore examined the allelic expression patterns in gonads and somatic organs of allotriploid (PAA) *S. alburnoides*, in which sequence differences resulting from the presence of distinct genomes can be used to follow expression of different alleles [5, 6]. Diagnostic restriction-fragment-length polymorphisms (RFLPs) and single-nucleotide polymorphisms (SNPs) between the A and P genomes were used to analyze the expression of three gonad-specific genes (*amh*, *dmrt1*, and *vasa*), a tissue-specific gene for the eye (*rhodopsin*), and the housekeeping genes *β-actin*, *rpl8*, and *gapdh*, ubiquitously expressed in all organs. RFLP analysis of *vasa* and *dmrt1* gene products in diploid hybrids (PA) revealed the presence of specific fragments, corresponding to both genomes (P and A). In triploids, a clearly distinct situation was observed (Figure S1): In most cases for *dmrt1*, and in all cases for *vasa*, the banding pattern indicated exclusive expression of the A genome allele. To confirm the observed pattern and to extend the analysis to all genes considered, we performed amplification and direct sequencing procedures for a total of 17 samples of diploid (4) and triploid (13) hybrids of *S. alburnoides*. Sequence analysis of cDNA polymorphisms of transcripts from gonads of the hybrid diploid fish revealed a biallelic expression of all genes. In triploids, three scenarios could be established, according to gene expression patterns in the gonad: (1) Four genes show expression exclusively from the A genome (*vasa*, β-*actin*, *rpl8*, and *gapdh*); (2) one gene shows a biallelic expression in all triploids (*amh*); and (3) one gene (*dmrt1*) shows biallelic expression in two individuals, whereas the remaining twelve express only the A genome (Figure 2A). The results were additionally confirmed by analysis of twenty independently obtained sequences of the four genes from two triploid individuals that showed differences in the allelic expression of *dmrt1* (PA versus AA).

We asked additionally whether this expression pattern was gonad specific or if genes in other organs could be subjected to similar allele-specific silencing. Polymorphisms at the cDNA level implied an overall biallelic expression of the eye-specific *rodopsin* gene in all diploids and triploids analyzed. In

*Correspondence: phch1@biozentrum.uni-wuerzburg.de

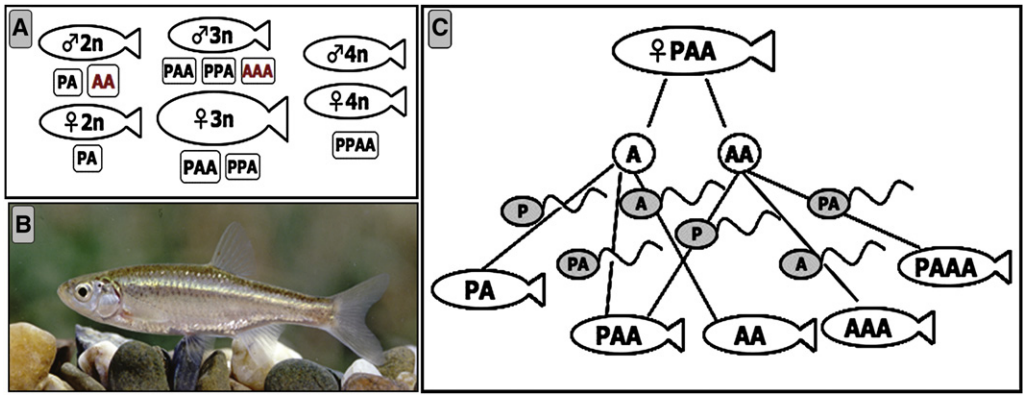


Figure 1. Diversity of Genotypes and Reproductive Dynamics of the *S. alburnoides* Complex

(A) Genotypes of *S. alburnoides* found in the Tejo River Basin, according to ploidy and sex. Genomes of *S. pyrenaicus* (maternal ancestor) and of the unknown paternal species are designated, respectively, as P and A.

(B) Specimen of *S. alburnoides* from the southern range of the complex (Tejo River Basin) (photo by Isabel Catalão).

(C) Reproductive mechanisms in the *S. alburnoides* complex of triploid PAA females: fertilization by P sperm, produced by normal meiosis by *S. pyrenaicus*; fertilization by A sperm, produced by normal meiosis by nuclear nonhybrid males; and fertilization by PA sperm, produced clonally by diploid or tetraploid hybrid males.

contrast, β-*actin* expression in liver, eye, and muscle of triploid *S. alburnoides* was in the majority of cases derived exclusively from the A genome(s) (Figure 2B). Exceptions were observed in two triploid individuals, for which β-*actin* in muscle and eye exhibited expression of the P and A genomes. Interestingly, for the same individuals, in the other organs, β-*actin* expression was exclusively due to the A genome(s). The same pattern of differential contribution of each genome was observed when *rp18* and *gapdh* expression in gonad and liver of triploid individuals was examined: Although all gonad samples were expressing only the A genome, the simultaneous occurrence of P and A transcripts was observed in some liver samples (Figure 3).

The identification of exclusive contribution of diagnostic alleles of the A genome in some organs and/or genes of triploid individuals by sequence analysis indicates that the P genome alleles would be inactivated in these samples. Considering the quantitative real-time PCR results, and assuming the most parsimonious hypothesis of silencing of only one allele in triploids, the observed dosage effects could result from the expression of the two homomorphic A alleles, indistinguishable both by quantitative and qualitative methods because of the absence of polymorphism. According to this hypothesis, exclusive expression of the P genome should not happen in triploid PAA hybrids, because this would result only in half the dose for each gene. In fact, no gene expression pattern corresponding exclusively to this genome was observed in our analysis. Considering the difficulties of determining the exact amount of mRNA in diploids compared to triploids, we cannot rule out the possibility that all three genome copies are

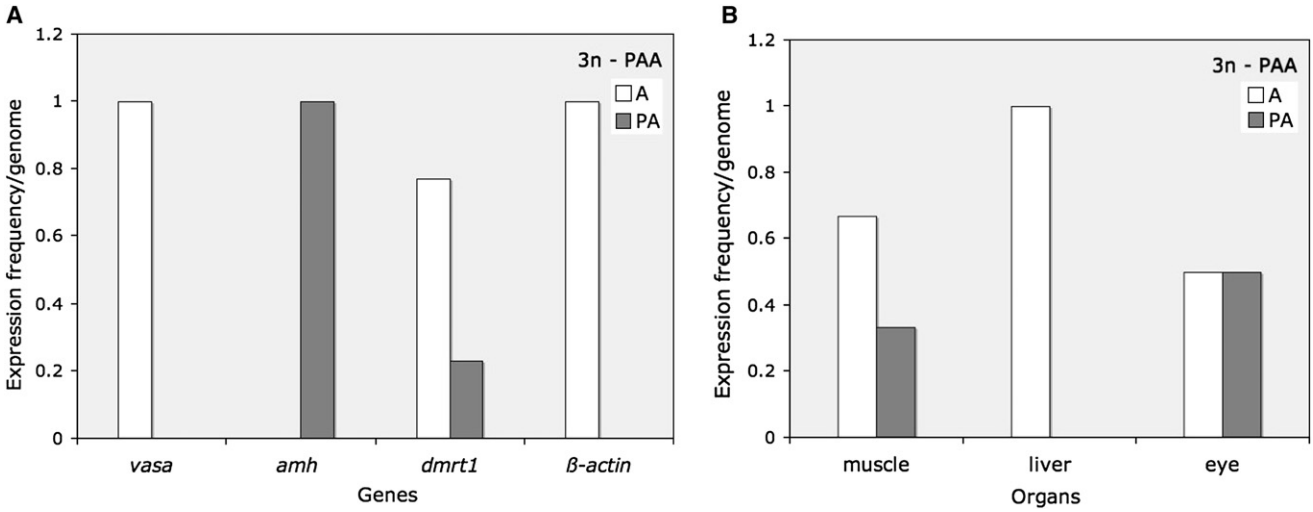


Figure 2. Frequency of Allele-Specific Transcripts in Different Organs

(A) Four genes (*vasa*, *amh*, *dmrt1*, and β-*actin*) in 13 gonad samples of triploid (PAA) *S. alburnoides* individuals.

(B) β-*actin* gene in different organs of triploid (PAA) *S. alburnoides* individuals.

Genome-specific polymorphisms were identified for each organ: muscle and liver over a total of ten individual samples; eye over four samples. The obtained frequencies of cDNA samples representing expression of the A, P, or both genomes are shown in the histograms.

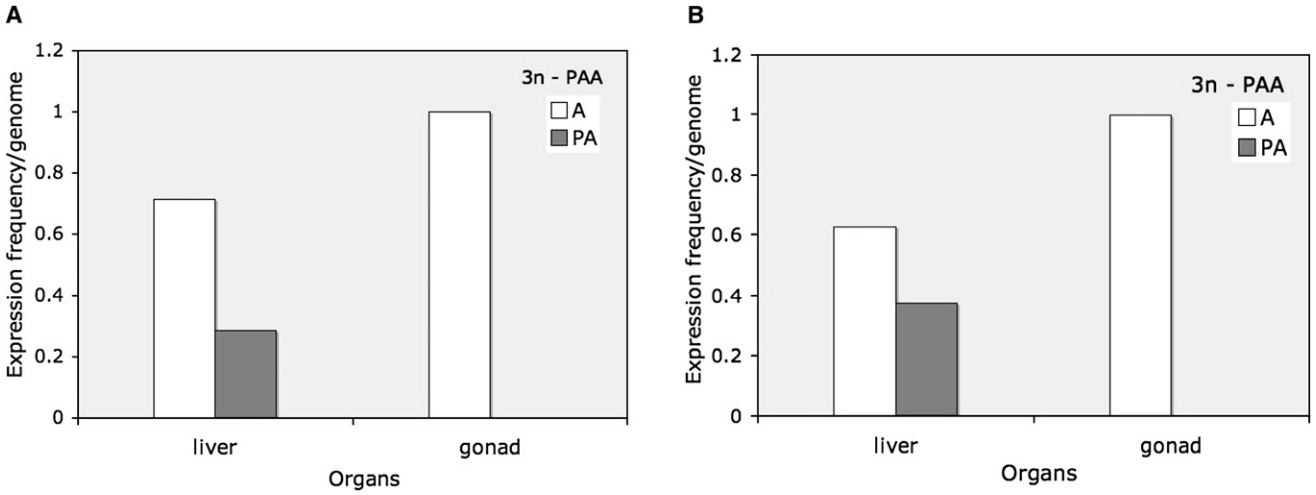


Figure 3. Frequency of Allele-Specific Transcripts in Different Organs

(A) *rpl8* and (B) *gapdh* genes in gonad and liver of triploid (PAA) *S. alburnoides* individuals. Genome-specific polymorphisms were identified for each organ: liver over a total of eight samples and gonad over a total of four samples. The obtained frequencies of cDNA samples representing expression of the A, P, or both genomes are shown in the histogram.

expressed in cases in which the P and A alleles were simultaneously identified. On the other hand, the hypothesis of silencing an entire haplome, more specifically, the silencing of the whole minority genome (P), as suggested by the exclusion of this genome from meiosis in triploid PAA *S. alburnoides* [5], was contradicted by these results, in which some genes show expression of A as well as P alleles. This hypothesis appeared initially to be a reasonable explanation because in a system where mechanisms account for genetic exclusion of a certain haplome (always P) as a way to overcome the "meiotic crisis" imposed by triploidy, the mechanism used to "mark" the P genome could also be used and operate in gene expression regulation.

A strong possibility to account for the observed gene-copy silencing would be the occurrence of genomic imprinting [9, 10], implying that gene expression control would be mediated specifically by one parental genome, whereas copies inherited from the other parent would be silenced in the hybrid. This idea would be supported by the possibility of occurrence of imprinting effects in fish, as demonstrated in inheritance of zebrafish transgenes [11]. However, our results do not seem to fit the expectations of a parental imprinting process [12]. In the case of a general imprinting mechanism, single-genome-mediated expression should be observed for the same gene in all organs in triploid hybrids, according to the parental origin of genomes. However, the actin gene analyzed here clearly showed organ-specific differences in allelic expression.

Thus, a less directional modulation mechanism should be considered to explain the maintenance of gene expression levels in *S. alburnoides*. Such a mechanism might operate in between the strict process of paternal origin-mediated expression, described for a number of vertebrates [13], and what more recently has been referred as a more plastic regulatory mechanism in plants [14]. One possibility to consider is random genome inactivation. In fact, the identification of PA alleles of *dmrt1* in a number of gonads and the differential expression of *β-actin* copies in somatic organs of the same individual could indicate that genome silencing would occur randomly. However, the results obtained were not only inconsistent, but clearly opposite to the 1:2 ratio of AA to PA, expected in case of random inactivation. The ratio of expressed PA copies was approximately 0.23 in *dmrt1* and 0.35 in *β-actin* ($p = 0.001$ for *dmrt1* and $p = 0.006$ for *β-actin*, calculated through 1000 randomizations and considering a null hypothesis of random inactivation of one of the genomes); this ratio was 0.29 in *rpl8* and 0.38 in *gapdh* ($p = 0.013$ for *rp18* and 0.075 for *gapdh*), thus implying an overall predominance of P genome-copy silencing for the genes analyzed.

Globally, the results fit a gene- rather than an organ- or haplome-specific silencing mechanism. The process does not seem to be completely random, because a clear preference in the expression of the majority (A) genome is observed. The means by which the specific contribution of each genome is regulated have yet to be determined, but it is possible that similar modulation mechanisms could be involved in the adjustment of gene expression in a diploid context as a consequence of differences in sex chromosome number [14] and in the coordination of gene dosage imbalance in more dramatically altered systems such as triploids. It remains an interesting point for future studies whether the phenomenon reported here acts without exception throughout the whole genome, if like in X chromosome inactivation a certain number of genes escape the dosage-regulation mechanism, or whether it is restricted to a specific subset of genes. As shown in recent studies, the levels of efficiency and stability of sex-chromosome compensation can vary between groups, ranging from a widespread silencing in humans to a more plastic process in fruit flies [15]. Likewise, it is possible that similarly a more plastic, individual-gene-mediated dosage-compensation mechanism also acts in hybrid fish. According to this, the observed exceptions for *dmrt1* and *β-actin* could be leaks in an ongoing evolution of a fine-tuning process regulating gene expression in response to overdosage.

The possibility of occurrence of this type of regulation brings up several questions not only regarding the functionality of gene dosage-compensation systems but also within the context of polyploid species evolution.

Here we report a gene-regulation mechanism involving gene-copy silencing in triploid *S. alburnoides*. Whether this

process might be exclusive to this complex is still to be determined, and divergent clues come from distinct groups. In the allotriploids of the genus *Poeciliopsis*, triallelic expression in a number of allozyme loci and the assessment of ploidy on the basis of allelic dosage suggest that all genome copies are contributing to overall expression and that no silencing occurs at the loci under analysis [16, 17]. Conversely, in other studies involving the triploid unisexual *Poecilia formosa*, indistinguishable expression between diploids and triploids was observed at several allozyme loci [18]. More recently, individual variation in pigmentation phenotypes resulting from crosses involving *Poecilia formosa* individuals indicates differential contribution of genomes to overall expression and suggests that a comparable gene-copy silencing process may also occur in this polyploid species [19]. Evidence of a mechanism that leads to a diploid state of genic activity also comes from comparisons between protein contents and enzyme activities of tetraploid and diploid counterparts of teleost [20, 21] and anuran [22] species. Thus, the same mechanism could also have a role in the regulation of gene expression in autopolyploids, but the possibility of a widespread occurrence among the different groups is yet to be demonstrated. Autopolyploid systems, however, offer less experimental potential in terms of understanding the gene-regulation process because, unlike allopolyploids such as *S. alburnoides* or *P. formosa*, the contribution of individual genomes cannot be identified on the basis of genome-specific sequence polymorphism.

In a broader sense, the present results imply that balanced expression and functional "diploidization" could in fact be necessary and that dosage compensation might be an extremely relevant factor contributing to the success and perpetuation of polyploids among lower vertebrates [23]. These findings might constitute a starting point toward the elucidation of important questions of genome evolution, namely, how genome regulation takes place in polyploid species, how dosage compensation can have a major role in the complex problem of odd genome regulation and perpetuation, and what factors allow lower vertebrates to endure and maintain ploidy changes so effectively.

Experimental Procedures

Samples

A total of 17 hybrid diploid (4) and triploid (13) specimens of *S. alburnoides* were collected from the Tejo River Basin (Portugal), where all forms that compose the complex are putatively present [7]. Five specimens of nuclear "nonhybrid" males of *S. alburnoides* and four of *S. pyrenaicus* (maternal ancestor of the complex) were morphologically identified and collected to be used as diploid controls of the A and P genomes. The determination of genotype identity of individual samples was performed by a conjoined approach of flow-cytometry measurements and the analysis of microsatellite variation.

Ploidy Determination

Blood samples were drawn from the caudal vein, stabilized in buffer (40 mM citric acid trisodium salt, 0.25 M sucrose, and 5% dimethyl sulfoxide), and immediately frozen at −80°C. Flow-cytometry measurements were conducted according to standard methods, used previously in the study of the complex [24].

DNA and RNA Extraction

Total genomic DNA was obtained from fin clips following standard protocols of digestion with SDS and proteinase K and phenol/chlorophorm extraction [25]. Total RNA was extracted from organs of *S. pyrenaicus* and *S. alburnoides* with the TRIZOL reagent (GIBCO-BRL) according to the supplier's recommendation. First-strand cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas) by use of oligo dT primers. Genomic DNA was additionally obtained from each organ sample used for RNA extraction.

Microsatellites

Individual genotypes were obtained by cross-species amplification of four microsatellite loci (LCO1, LCO3, LCO4, and LCO5) and genome-specific allele identification, as already used in the study of the complex [26]. Diploid *S. alburnoides* were genotyped as PA (hybrids) and AA (nuclear "nonhybrid"). All triploids were genotyped as PAA from fin clips and identified by the presence of three alleles for at least one locus. Organ genotypes in each individual were checked by parallel amplification of the four microsatellite loci.

Gene Isolation

Sequences of the teleost orthologs of a total of seven genes, both organ-specific (*amh*, *dmtr1*, *vasa*, and *rodopsin*) and ubiquitously expressed (*β-actin*, *rpl8*, and *gapdh*), were used as templates for the design of degenerated primers for each candidate gene. Forward primers were tested in rapid amplification of cDNA ends. First-strand cDNA (including the 3′-end information of mRNA) was reversed transcribed from testis RNA with the 3′ RACE System (Invitrogen). Amplification was performed with an adaptor primer and the gene-specific forward primer. PCR conditions were as follows: preheating at 94°C for 3 min, 30 cycles at 94°C for 45 s, 52°C for 40 s, and 72°C for 3 min and a final extension at 72°C for 15 min. Under the same conditions, primer pairs were tested in *S. pyrenaicus* (PP) and *S. alburnoides* (AA) cDNA samples, and the resulting sequences from the two procedures were used as templates for designing specific primers (Table S2). Polymorphic sites for the two genomes (A and P) were identified for the seven genes by sequence alignment, with Sequencher ver. 4.0 (Gene Codes Corporation).

Sequence Analysis and Genome Expression

cDNA samples from adult gonad of diploid (PA) and triploid (PAA) hybrid *S. alburnoides* were used as templates for independent amplification and direct sequencing of gene products of six genes: *amh*, *dmrt1*, *vasa*, *rpl8*, *gapdh*, and *β-actin*. On the basis of the identified polymorphic sites between the two genomes, the presence of single-genome copies (exclusively A or P) or of both genomes (identified in ambiguous positions) was determined through sequence comparison. Two (*vasa*, *rp18*, *gapdh*) to four sequences (*dmrt1*, *amh*, *β-actin*) were obtained per individual per organ. For two triploid individuals (genotyped as PAA), a total of 20 sequences for each gene were obtained, from two independent RNA extractions, four distinct cDNA-synthesis procedures, and independent PCR reactions. Partial sequences of the *β-actin*, *rpl8*, *gapdh*, and *rodopsin* genes were obtained from organ samples (liver, muscle, and in the latter case, only eye) of diploid and triploid hybrid individuals. According to the presence of sequence differences at informative positions, the expression of the two or only one genome type was determined for the obtained sequences. The significance of the observed frequencies of expression was assessed by 1000 randomizations, considering the null hypothesis of random inactivation of each genome.

Restriction Analysis

Partial gene products, obtained from cDNA amplification with gene-specific primers, were digested with enzymes that generate differential restriction-fragment polymorphisms between the P and A genomes, identified by MetaPhor Agarose (Cambrex) gel electrophoresis. AciI was used for the differential restriction of *amh* and *dmrt1* gene products, SfaNI for *β-actin*, RsaI for *rodopsin*, and BfaI for *vasa*. Restriction was performed for 1 hr at 37°C. For an example, see Figure S1.

Determination of β-Actin-RNA/DNA Levels

To assess whether there was a decrease in the RNA amount relative to genomic DNA content in triploids, we have conducted PCR experiments for the single-copy *β-actin* gene, with DNA and cDNA samples from liver of three diploid *S. pyrenaicus* and two triploid *S. alburnoides* as templates. Genomic DNA and total RNA were extracted simultaneously from each sample with the AllPrep DNA/RNA Mini Kit (QIAGEN). Amplification of DNA and cDNA templates was conducted with *β-actin* primers (Table S3). PCR conditions were as follows: preheating at 94°C for 3 min, 30 cycles at 94°C for 45 s, 60°C for 40 s, and 72°C for 3 min and a final extension at 72°C for 15 min. Products were run in a 1.5% agarose gel and densitometrically quantified with ImageG (Figure S2).

Real-Time PCR Analysis

Primers and specific TaqMan probes for the *β-actin*, *rpl8*, *gapdh*, *dmrt1*, *amh*, and *vasa* genes were designed on the basis of the common regions of sequences of the P and A genomes (obtained from *S. pyrenaicus* and *S. alburnoides* samples) (Table S3). Reactions were conducted on an iCycler iQ Real Time PCR equipment (50°C for 2 min; 95°C for 10 min; 50 cycles at 60°C for 1 min), visualized with iCycler iQ Real-Time Detection System Software. A subsample of muscle, liver, eye, and gonad RNA of individuals of different sex, genotype, and ploidy were selected for amplification. Total RNA quantification was performed with a Nanodrop ND-1000 Spectrophotometer, and normalization of template RNA quantity was performed before cDNA synthesis with SUPERSCRIPT III 1st Strand cDNA Synthesis Kit (Invitrogen), with oligo dT primers. Relative quantification was performed against initial quantity of nucleic acid, and reaction efficiency and replicability was assessed by generation of a standard curve. Triplicates of each sample were used both for standard curve generation and during experimental assays. Amplification of the same samples was performed three times, and reproducibility of results was confirmed. Relative expression ratios were determined through direct comparison of Ct value averages of triplicates between diploid controls (AA, PP, and PA) and triploid (PAA) samples, with the approach described in a study of Z-linked gene expression and sex-specific dosage compensation in birds [27].

Supplemental Data

Supplemental Data include three tables and two figures and can be found with this article online at http://www.current-biology.com/cgi/content/full/18/17/1344/DC1/.

Acknowledgments

We thank Miguel Brito and Joana Vacas for assistance and expertise in the quantitative real-time PCR experiments. This research was supported by a grant from the Fundação para a Ciência e a Tecnologia (FCT; part of the Ministry of Science of Portugal) to I.P. (SFRH/BD/19286/2004), by the Project PTDC/BIA-BDE/65586/2006 (FEDER participation), and by the project C-5 funded by the SFB 487 of the Deutsche Forschungsgemeinschaft.

Received: April 25, 2008
Revised: July 18, 2008
Accepted: July 18, 2008
Published online: September 4, 2008

References

1. Antonarakis, S.E., Lyle, R., Dermitzakis, E.T., Reymond, A., and Deutsch, S. (2004). Chromosome 21 and down syndrome: From genomics to pathophysiology. Nat. Rev. Genet. *5*, 725–738.
2. Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. Annu. Rev. Genet. *34*, 401–437.
3. Dowling, T.E., and Secor, C.L. (1997). The role of hybridization and introgression in the diversification of animals. Annu. Rev. Ecol. Syst. *28*, 593–620.
4. Vrijenhoek, R.C. (1994). Unisexual fish: Model systems for studying ecology and evolution. Annu. Rev. Ecol. Syst. *25*, 71–96.
5. Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H., and Birchler, J.A. (2005). Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics *169*, 389–397.
6. Suzuki, M.G., Shimada, T., Yokoyama, T., and Kobayashi, M. (1999). The influence of triploidy on gene expression in the silkworm, *Bombyx mori*. Heredity *82*, 661–667.
7. Alves, M.J., Coelho, M.M., and Collares-Pereira, M.J. (2001). Evolution in action through hybridization and polyploidy in an Iberian freshwater fish: A genetic review. Genetica *111*, 375–385.
8. Alves, M.J., Coelho, M.M., and Collares-Pereira, M.J. (1998). Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): A way to avoid the genetic constraints of uniparentalism. Mol. Biol. Evol. *15*, 1233–1242.
9. Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development *125*, 3329–3341.
10. Alleman, M., and Doctor, J. (2000). Genomic imprinting in plants: Observations and evolutionary implications. Plant Mol. Biol. *43*, 147–161.
11. Martin, C.C., and McGowan, R. (1995). Parent-of-origin specific effects on the methylation of a transgene in the zebrafish, *Danio rerio*. Dev. Genet. *17*, 233–239.
12. McGowan, R.A., and Martin, C.C. (1997). DNA methylation and genome imprinting in the zebrafish, *Danio rerio*: Some evolutionary ramifications. Biochem. Cell Biol. *75*, 499–506.
13. Lucchesi, J.C. (1978). Gene dosage compensation and the evolution of sex chromosomes. Science *202*, 711–716.
14. Birchler, J.A., Riddle, N.C., Auger, D.L., and Veitia, R.A. (2005). Dosage balance in gene regulation: Biological implications. Trends Genet. *21*, 219–226.
15. Straub, T., and Becker, P.B. (2007). Dosage compensation: The beginning and end of generalization. Nat. Rev. Genet. *8*, 47–57.
16. Quattro, J.M., Avise, J.C., and Vrijenhoek, R.C. (1992). Mode of origin and sources of genotypic diversity in triploid gynogenetic fish clones (Poeciliopsis: Poeciliidae). Genetics *130*, 621–628.
17. Mateos, M., and Vrijenhoek, R.C. (2005). Independent origins of allotriploidy in the fish genus *Poecilliopsis*. J. Hered. *96*, 32–39.
18. Turner, B.J., Balsano, J.S., Paul, J.M., and Rasch, E.M. (1983). Clonal diversity and evolutionary dynamics in a diploid-triploid breeding complex of unisexual fishes (Poecilia). Evolution Int. J. Org. Evolution *37*, 798–809.
19. Lamatsch, D.K., Fischer, P., Lampert, K.P., Geiger, M., Schlupp, I., and Schartl, M. (2008). Diploid Amazon mollies (Poecilia formosa) show a higher fitness than triploids in clonal competition experiments. Evol. Ecol., in press. Published online May 20, 2008. 10.1007/s10682-008-9264-2.
20. Schmidtke, J., and Engel, W. (1975). Gene action in fish of tetraploid origin. I. Cellular and biochemical parameters in cyprinid fish. Biochem. Genet. *13*, 45–51.
21. Schmidtke, J., and Engel, W. (1976). Gene action in fish of tetraploid origin. V. Cellular RNA and protein content and enzyme activities in cyprinid, clupeoid and salmonid species. Biochem. Genet. *14*, 975–980.
22. Beqak, W., and Pueyo, M.T. (1970). Gene regulation in the polyploid amphibian *Odonto-phrynus americanus*. Exp. Cell Res. *63*, 448–451.
23. Otto, S.P. (2007). The evolutionary consequences of polyploidy. Cell *131*, 452–462.
24. Dawley, R.M., and Goddard, K.A. (1988). Diploid-triploid mosaics among unisexual hybrids of the minnows *Phoxinus neogaeus*. Evolution Int. J. Org. Evolution *42*, 649–659.
25. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215.
26. Crespo-López, M.E., Pala, I., Duarte, T., Dowling, T.E., and Coelho, M.M. (2007). Genetic structure of the diploid–polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. J. Fish Biol. *71*, 423–436.
27. McQueen, H.A., McBride, D., Miele, G., Bird, A.P., and Clinton, M. (2001). Dosage compensation in birds. Curr. Biol. *11*, 253–257.

CHAPTER 4 | Supplementary data

Supplemental Data

Dosage Compensation by Gene-Copy Silencing in a Triploid Hybrid Fish

Irene Pala, Maria M. Coelho, Manfred Schartl

Table S1. Relative expression ratio between diploid and triploid organ samples for the *β actin, rpl8, gapdh*, *vasa*, *dmrt1* and *amh* genes. Genotype (G); sample identification number (#); Ratio between triploid:diploid (control) average threshold cycle (Ct) values [reference control: (**a**) Sample:AA; (**b**) Sample:PP; (**c**) Sample:PA]; Average ratios for each genotype, relative to diploid expression values, according to gene and organ (consistent with an 1:1 ratio between triploid and diploid samples).

GENE	ORGAN	G	ID#	EXPRESSION Av Ct ratio	Av ratio/ genotype
β actin					
	Muscle	PAA	87	0.956 (a)	
			102	1.032 (a)	
			115	0.949 (a)	
					0.979 (a)
	Liver	PAA	87	0.901 (c)	
			102	1.051 (a); 1.022(c)	
			105	0.955 (a); 0.923(c)	
			108	1.036 (a); 0.960(c)	
			112	0.990(a); 0.917 (c)	
			115	0.957 (c)	
			117	1.068 (a); 0.996(c)	
					1.020(a); 0.954(c)
	Eye	PA	85	1.046 (a); 1.076(b)	
			86	1.042(a); 1.074(b)	
					1.044(a); 1.076(b)
		PAA	101	0.974(a); 1.004(b)	
			102	1.062(b)	
			105	0.997(a); 1.050(b)	
			118	0.997 (a); 1.027(b)	
					0.989(a); 1.036(b)
	Gonad	PAA	101	1.016(b)	
			102	0.983(a); 1.003(b)	
			105	1.000(a); 1.023(b)	
			118	0.996(a); 0.998(b)	
					0.993(a); 1.010(b)

rpl8					
	Liver	PAA	87	0.988(a)	
			88	1.070(a)	
			108	1.218(a)	
			112	1.072(a)	
			115	1.211(a)	
					1.112(a)
	Eye	PA	85	0.922(a)	
					0.922(a)
		PAA	88	1.020(a); 1.106(c)	
			101	0.894(a); 0.969(c)	
			102	0.829(a); 0.899(c)	
			108	0.902(a); 0.978(c)	
					0.911(a); 0.988(c)
	Gonad	PAA	88	0.872(a); 0.827(b)	
			89	0.935(a); 0.887(b)	
			101	0.968(a); 0.918(b)	
			102	0.877(a); 0.832(b)	
					0.913(a); 0.866(b)
gapdh					
	Liver	PAA	87	0.981(a)	
			88	1.138(a)	
			108	1.185(a)	
			112	1.185(a)	
			115	1.098(a)	
					1.118(a)
	Eye	PAA	88	1.095(a)	
			101	1.046(a)	
			102	1.065(a)	
			108	1.090(a)	
					1.074(a)
	Gonad	PAA	88	1.049(b)	
			89	1.082(b)	
			101	1.167(b)	
			102	1.038(b)	
					1.084(b)
vasa					
	Gonad	PAA	89	0.913(a); 1.019(b)	
			101	1.085(a); 1.163(b)	
			102	1.063(a); 1.087(b)	
			105	1.189(a); 1.231(b)	
					1.063(a); 1.125(b)
dmrt1					
	Gonad	PAA	89	0.926 (a); 0.976(b)	
			101	0.943(a); 0.991(b)	
			102	0.944(a); 0.994(b)	
			105	0.994(a); 1.046(b)	

				0.952(a); 1.002(b)
amh				
	Gonad	PAA	89	1.146(a); 1.006(b)
			101	1.127(a); 1.133(b)
			102	1.064(a); 1.195(b)
			105	1.135(a); 1.274(b)
				1.118(a); 1.152(b)

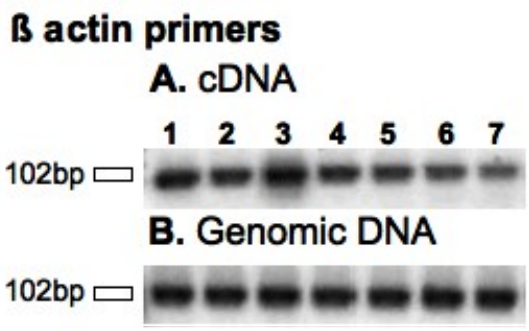
Table S2. Primer sequences (primers designed on the present work), references and GenBank accession numbers for each gene (for the P and A genomes).

Gene	**Primer**	**Sequence/ Ref**	**Accession no.**
amh	AMH- F2	(S1)	EU136185
	AMH-R2		EU136186
vasa	VASA- F	(S2)	EU199437
	VASA- R1 D	5'-TGCTCCACATCACTGCA-3'	EU199438
dmrt1	DMRT1- F1 D	5'-ATGAGYGARGARGAGCA-3'	EU199439
	DMRT1 zf-R3	5'-GGCTGGTAAAGGTTGTAATAG-3'	EU199440
ß-actin	ß-ACTIN-F1	5'-CAACGGCTCCGGCATGTG -3'	EU199435
	ß-ACTIN-R1	5'-TGCCAGGGTACATGGTGG-3'	EU199436
rhodopsin	Rho_F	(S3)	EU199442
	Rho_R		EU199441
rpl8	rpl8 Forward	(S4)	EU542915
	rpl8 Reverse		EU542916
gapdh	GAPDH F1	5'- ATCAGGCATAATGGTTAAAGTTGG-3'	EU542913
	GAPDH R1	5'- AACTCATTGTCATACCATGTGACC- 3'	EU542914

Table S3. Specific primers and TaqMan probes based on *amh*, *vasa*, *dmrt1* and *ß-actin* sequences of *S. pyrenaicus* and *S. alburnoides*.

Primer	**Sequence**
AMH real-F1	5'- AGGCCACWCGTTTCCTGATTC - 3'
AMH real-R1	5'- TTCGAGCCRCTTTSATCCTTCC - 3'
VASA real-F1	5'- GCCAGGAAGTTCGCATTTGG - 3'
VASA real-R1	5'- CAACACCTCTCGAATAGTGTATCC - 3'
DMRT1 real-F1	5'- CATCACCCACCAACAGCAGT- 3'
DMRT1 real-R1	5'- GTAATAGGAGGCATCCACCATCA- 3'
ß-ACTIN real-F1	5'- CCACACCGTGCCCATCTATG- 3'
ß-ACTIN real-R1	5'- GGTCAGGATCTTCATCAGGTAGTC - 3'
RPl8 real-F1	5'- GACCAAGAAATCCAGAGTCAAGC- 3'
RPl8 real-R1	5'- TGCCTTCAGGATGGGTTTGTC- 3'
GAPDH real-F1	5'- GTTGGTCATCGAYGGTCACG- 3'
GAPDH real-R1	5'- ACGTATGTAGCACCTGCATCAC- 3'
Probe/ Label	**Sequence**
AMH / 6-FAM	5'- CGGGAGGACACAGTCATGCCCACG - 3'
VASA/ 6-FAM	5'- ACCGCCATACACCACCACAGGACG- 3'
DMRT1/ 6-FAM	5'- ACCGCATCTCCCATAGAGAGCCGT - 3'
ß-ACTIN/ VIC	5'- CGACCAGCCAGATCCAGACGCAGG - 3'
RPl8/ 6-FAM	ACGACCACCGCCAGCAACAACACC
GAPDH/ 6-FAM	TCTTCAGCGAGAGGGACCCAGCCA

a.



b.

Sample/G	Within sample cDNA/DNA ratio
1/ PP	0.90
2/ PP	0.90
3/ PP	1.19
4/ PP	0.80
5/ PAA	**0.71**
6/ PAA	**0.64**
7/ PAA	**0.59**

Figure S1. **(a)** Amplification products of the single copy ß-actin gene, using cDNA and genomic DNA from the same tissue sample as template– samples 1 to 4: *S. pyrenaicus* of PP genotype; samples 5 to 7: *S. alburnoides* of PAA genotype**. (b)** Within sample ratio of densitometrically determined amounts of cDNA/ DNA PCR products for all ploidy/ genotype (G) combinations – values obtained for triploids are indicated in bold (expected ratio according to a diploid dosage compensation in triploid samples, would be 0.67).

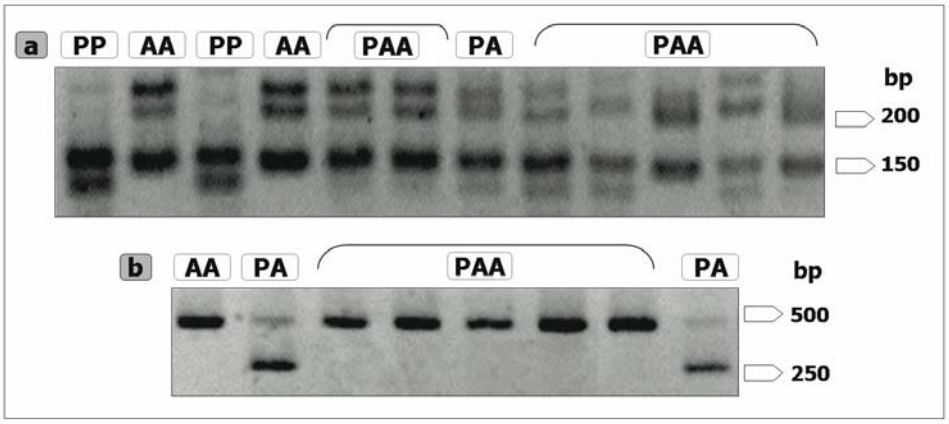


Figure S2. Example of restriction patterns obtained from the digestion with AciI and BfaI of *S. alburnoides* (nuclear non-hybrid –AA; diploid and triploid hybrid samples – PA and PAA) and *S. pyrenaicus* (PP) of RT-PCR products of *dmrt1* (**a**) and *vasa* (**b**) transcripts. Sample genotypes are indicated above lanes.

References

S1. Pala I, Klüver N, Thorsteinsdóttir S, Schartl M, Coelho MM (2008) Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*. Gene 410: 249-258.

S2. Ma C, Fan L, Ganassin R, Bols N, Collodi P (2001) Production of zebrafish germ-line chimeras from embryo cell cultures. Proc Natl Acad Sci U S A 98: 2461-2466.

S3. Fitzgibbon J, Hope A, Slobodyanyuk SJ, Bellingham J, Bowmaker JK, Hunt DM (1995) The rhodopsin-encoding gene of bony fish lacks introns. Gene 164: 273-277.

S4. Filby AL, Tyler CR (2007) Appropriate "housekeeping" genes for use in expression profiling the effects of environmental estrogens in fish. BMC Mol Biol 8: 8-10.

CHAPTER 5

Gene expression regulation and lineage evolution: the North and South tale of the *Squalius alburnoides* complex

Irene Pala, Manfred Schartl, Maria Manuela Coelho

In prep

Gene expression regulation and lineage evolution: The North and South tale of the *Squalius alburnoides* complex

Irene Pala [a], Manfred Schartl [b], Maria Manuela Coelho [a]

[a] Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

[b] Physiologische Chemie I, Biozentrum, University of Würzburg, Würzburg, Germany

Abstract

The evolution of hybrid polyploids, their viability and perpetuation over evolutionary time have always been questions of great interest, and several plant and animal polyploid groups have been extensively studied in that respect. In hybrid polyploid vertebrates little is known about the gene regulatory impacts of hybridisation and polyploidisation events. The *Squalius alburnoides* complex of hybrid fish is a very attractive system to address these questions, as it includes a wide variety of diploid and polyploid forms and intricate systems of genetic exchange and diversity maintenance. The complex has been regarded as an example of success among polyploid vertebrates, but significant differences have been found between populations, regarding genetic diversity, viability and evolutionary potential.

Through the study of the patterns of genome specific allele expression in seven housekeeping and tissue specific genes, in three main river basins of *S. alburnoides* distribution, we have shown that gene expression patterns vary within the complex, according to geographical origin and the type of genome involved in the hybridisation process. In polyploid individuals of the southern populations of *S. alburnoides* there is an overall tendency for the silencing of the heteromorphic genome gene copies, while in northern populations both triploids and tetraploids exhibit biallelic gene expression patterns in most genes analysed, irrespective of genomic composition. Based on the observed patterns, the history and the possibilities of evolution of each population we propose that different evolutionary pathways might be followed by distinct lineages even within the same polyploid complex and that gene expression might also vary accordingly, depending on particular genome contexts. The present findings may provide initial clues towards the understanding of the pathways that lead to hybrid lineage progression and offer evidence to a broader perspective on the possibilities of vertebrate polyploid evolution.

Keywords: allopolyploid; gene expression; allele silencing; *S. alburnoides*; hybrid lineage evolution

1. Introduction

Ploidy rise brought upon by a hybridisation event, creates an epigenetic instability state that can only be overcome if regulation mechanisms that contribute to gene copy perpetuation and heterosis outrun disadvantages and potentiate species adaptation and viability.

Hybrid polyploidy success has been extensively reported in plant species (Otto & Whitton, 2000) and more restrictedly in animals (Dowling & Secor, 1997), but the regulatory changes that contribute to genome stabilization and regulation in the presence of distinct chromosome sets are still elusive. Successful polyploid vertebrates have been extensively studied regarding different molecular aspects that contribute to their viability. Dynamics of genome interaction and heterosis have been elucidated in *Rana esculenta* (Hotz *et al.*, 1999) and particular reproductive and chromosomal dynamics have been reported in triploids of the *Bufo viridis* complex (Stöck *et al.*, 2002). The molecular aspects of perpetuation and diversification of distinct genomes brought together in an asexual context have been widely characterized in the Amazon molly, *Poecilia formosa* (Lampert *et al.*, 2007). In all these systems, aspects such as the relevance of heterosis and the role of gene duplicates in lineage specific evolution have been comprehensively addressed (Comai, 2005) but the impact of polyploidisation on the mechanisms regulating gene expression have not yet been clarified. In plants, it has been shown that changes in gene expression and the phenotypic consequences of polyploidy should have important consequences on the evolutionary potential of polyploids (Pignatta *et al.*, 2008). Triploids are a particularly complex evolutionary jigsaw. Several deviations in the expected gene expression patterns have been reported in polyploid plants (Auger *et al.*, 2005), including organ-specific silencing (Adams *et al.*, 2003), epigenetic regulation of duplicates (Comai *et al.*, 2000; Shaked *et al.*, 2001), parent-of-origin-specific control of gene expression (Alleman & Doctor, 2000) and a still unclear effect designated as odd ploidy response (Guo *et al.*, 1996). In vertebrates, the effects of uneven ploidy on gene regulation and their impact in the evolutionary potential of populations have seldom been addressed. Polyploidy has been proposed as an important driving force of evolution and the success of polyploid lineages confirms it as a successful evolutionary transition and a potentially relevant factor in evolutionary diversification (Otto & Whitton, 2000). Therefore, it would be important to investigate the functional basis of how polyploids overcome an initial period of instability and establish processes that allow evolutionary flexibility and efficient competition with their diploid counterparts.

The *Squalius alburnoides* complex of hybrid fish presents several unique features among polyploid vertebrates that make it a perfect system to address the question of gene expression regulation and its impact on the persistence of hybrid lineages over evolutionary time (reviewed in Alves *et al.*, 2001). It is an allopolyploid, resulting from interspecific hybridisation between two Iberian species:

Squalius pyrenaicus as the maternal ancestor (P genome) and a still undetermined species, closely related to *Anaecypris hispanica* as the paternal ancestor (A genome) (reviewed in Alves *et al.*, 2001). Several scenarios have been proposed regarding the time and location of the origin of the complex. Five independent origins of the complex have been postulated, based on the analysis of the complete cytochrome b gene, dating the hybridisation in an extended time period, going back to the upper Pliocene (Cunha *et al.*, 2004). In the northern region of the Iberian Peninsula (Mondego and Douro River basins), the complex should have been established by local hybridisation with *S. pyrenaicus*, with the latter species being subsequently wiped out from these locations by vicariant events (Cunha *et al.*, 2004). Alternative hypotheses of single origin followed by dispersion have also been proposed (Alves et. al, 1997; Sousa-Santos *et al.*, 2007). It has been postulated that the complex might have originated in the area of what is now the Guadiana River, less than 0.7 MY ago, and subsequently dispersed through the connections between river basins. According to this scenario, the Mondego and Douro Basins would have been colonized from Tejo, respectively at around 0.05 MY and 0.01 MY ago, through stream capture involving adjacent tributaries (Sousa-Santos *et al.*, 2007). In any case, all authors provided evidence of heterogeneity within the distribution of the complex and the occurrence of lineages that were differentially established in time. Globally, the complex has been regarded as an example of success among polyploid vertebrates, but differences in the genetic diversity, viability and evolutionary potential emerged when analyzing populations of the complex from different geographical locations.

Presently, the complex is widely distributed in the Iberian Peninsula where it occurs in sympatry with two species of the genus Squalius: *S. pyrenaicus* in southern basins and *Squalius carolitertii* (C genome) in northern drainages. These bisexual species interact directly with the complex, acting as sources of genetic material and contributing to the maintenance of gene flow and the continuous cycling between forms (reviewed in Alves *et al.*, 2001). The complex is composed of animals with different ploidy degrees and genomic constitutions, including diploids (PA, CA), triploids (PAA, PPA, CAA and CCA) and tetraploids, with C or P genome inclusion according to geographical location. In southern populations an additional form, designated as “nuclear non hybrid” (AA), occurs and actively contributes to the maintenance of the genetic diversity of the complex (Alves *et al.*, 1999). All forms are apparently fertile and interact through diverse reproductive modes that include rare gynogenesis and the more common processes of hybridogenesis and meiotic hybridogenesis, in which one genome is discarded and the others undergo “normal” recombination and meiosis (Carmona *et al.*, 1997; Alves *et al.*, 1998, 2004; Pala & Coelho, 2005). The escape from strictly asexual reproductive modes and the complexity of the genetic exchange routes contribute

definitively to the success of *S. alburnoides* and for its unique position as an example of viability among polyploid taxa.

The study of such a system would be highly valuable as an initial approach to the question of gene expression regulation and the evolution of polyploid taxa. As an allopolyploid and due to the presence of different genomes, the *S. alburnoides* complex offers the advantage of allowing the distinction between the different genome-specific gene copies contributing to overall expression. Thus we can qualitatively identify which genomes are being expressed. Additionally, the presence of lineages within the complex, established differentially in time, including different forms and showing distinct dynamics of genetic exchange and different evolutionary potential, further allow us to explore the question of how regulation patterns could play a relevant role in lineage evolution and persistence upon a similar hybridisation event.

We have studied the expression pattern of a total of seven genes, both tissue-specific and widely expressed in the organism in different organs of diploid, triploid and tetraploid forms of northern and southern populations of the complex. In the present work we report differential expression patterns of genome-specific gene copies according to geographical location. We have additionally compared the differential expression patterns with the selective forces acting on the different genomes brought together by hybridisation and the evolutionary history of the different populations. Based on our data and the present knowledge of the complex, we have proposed several hypotheses that could relate mechanisms of gene expression regulation brought on upon hybridisation and ploidy rise, to the dynamics of genome evolution and polyploid lineage perpetuation.

2. Methods

2.1. Samples

Samples of *S. alburnoides* and of the bisexual sympatric *Squalius* species were collected from three locations corresponding to the northern and southern distribution areas of the complex: the Douro and the Mondego River Basins in which it occurs in sympatry with *S. carolitertii*; and the Tejo River Basin in which it occurs in sympatry with *S. pyrenaicus* (Fig. 1). A total of ten hybrid specimens of *S. alburnoides* and two specimens of *S. carolitertii* were collected from Paiva River, a tributary of the Douro Basin. Five hybrid *S. alburnoides* and two *S. carolitertii* were collected from the Ceira River, a tributary of the Mondego Basin. Twenty *S. alburnoides* hybrids and two sympatric *S. pyrenaicus* were collected from Sorraia River within the Tejo River Basin. Individuals were sacrificed with an overdose of anaesthetic MS222 and organs were collected for RNA and DNA extraction and stored at -80°C.

Fin clips were collected from each individual, stored in ethanol at 4°C and were later used for DNA extraction. The genotype identity of individual samples was determined by flow cytometry measurements and the analysis of microsatellite variation.

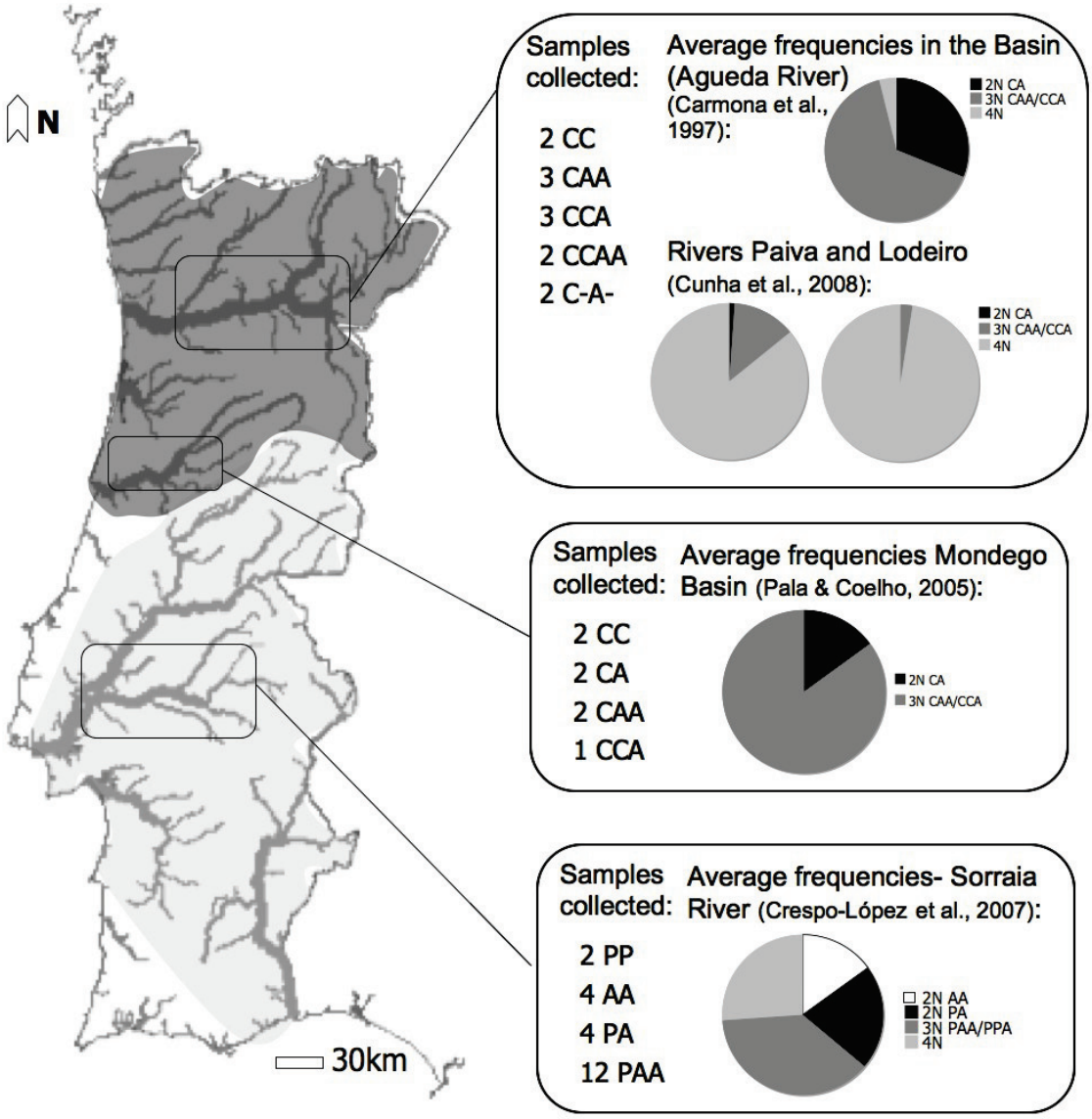


Figure 1. Distribution of *S. alburnoides* in the areas of sympatry with *S. carolitertii* (dark grey) and *S. pyrenaicus* (light grey), composition of samples collected in each location (Douro, Mondego and Tejo River basins) and reported frequencies of diploid (2N), triploid (3N) and tetraploid (4N) forms in the same River basin.

2.2. Ploidy determination.

Blood samples were drawn from the caudal vein, stabilized in buffer (40 mM citric acid trisodium salt, 0.25 M sucrose, and 5% dimethyl sulfoxide) and immediately frozen at -80°C. Flow cytometry measurements were conducted according to standard methods (Dawley & Goddard, 1988).

2.3. DNA and RNA extraction.

Total genomic DNA was obtained from fin clips following standard protocols of digestion with SDS and proteinase K and phenol/chlorophorm (Miller *et al.*, 1988). Total RNA was extracted from organs of *S. pyrenaicus*, *S. carolitertii* and *S. alburnoides* using the TRIZOL reagent (Gibco-BRL) according to the supplier's recommendation. First strand cDNA was synthesized with RevertAid™ First Strand cDNA Synthesis Kit (Fermentas) using oligo dT primers. Genomic DNA was additionally obtained from each organ sample used for RNA extraction.

2.4. Microsatellites.

Individual genotypes were obtained by cross species amplification of three microsatellite loci (LCO3, LCO4 and LCO5) and genome specific allele identification, as previously performed in the study of the complex (Pala & Coelho, 2005; Crespo-López *et al.*, 2007; Cunha *et al.*, 2008). Reactions were conducted in group, using the Multiplex PCR Kit (Qiagen). PCR conditions were according to the supplier's recommendation, except for the annealing temperature, which was lowered to 52°C. Reactions were performed using reverse primers labelled with fluorophore and amplification products were analysed with an automated sequencer (ABI 310 Genetic Analyser). In the cases for which it was not possible to identify the three or four distinct alleles in triploids and tetraploids, an additional locus, LCO1 was used to determine the complete genotype. Individual organ genotypes were checked by parallel amplification of the microsatellite loci.

2.5. Gene isolation.

Sequences of the teleost orthologs of a total of seven genes, gonad specific (*amh*, *dmrt1*), eye specific (*rhodopsin*), and ubiquitously expressed (*ß-actin*, *rpl8, ef1a* and *gapdh*), were used as templates for the design of gene-specific primers (Table I). Primers were initially tested in cDNA samples of *S. pyrenaicus* (PP), *S. carolitertii* (CC) and *S. alburnoides* (AA). Polymorphic sites for the three genomes (P, C and A) were identified for the seven genes by sequence alignment, using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.). Conserved genome-specific polymorphisms identified in "pure genome" samples were used as reference for the analysis of sequences obtained from hybrid samples. Organ samples of diploid (PA, CA), triploid (PAA, CAA, CCA) and tetraploid

(CCAA, C-A-) *S. aburnoides* were amplified according to the following PCR conditions: pre-heating at 94°C for 2 min 30 s, 35 cycles at 94°C for 45 s, 52°C (*amh*, *dmrt1*)/ 55°C (*rhodopsin, ß-actin*, *rpl8, ef1a* and *gapdh*) for 40 s and 72°C for 1 min 15 s and a final extension at 72°C for 10 min.

Table I. Primer sequences (primers designed in the present work), references and GenBank accession numbers for each gene (diagnostic alleles for the P, A and C genomes).

Gene	Primer	Sequence/ Ref	Genome	Accession nr.
amh	AMH-F2 AMH-R2	Pala *et al.*, 2008a	P A C	EU136185 EU136186 FJ587489
dmrt1	DMRT1- F1 D DMRT1 zf-R3	Pala *et al.*, 2008b	P A C	EU199439 EU199440 FJ587490
rhodopsin	Rho-F Rho-R	Fitzgibbon *et al.*, 1995	P A	EU199442 EU199441
ß-actin	ß-ACTIN-F1 ß-ACTIN-R1	Pala *et al.*, 2008b	P A C	EU199435 EU199436 FJ587491
rpl8	Rpl8 forward Rpl8 reverse	Pala *et al.*, 2008b	P A C	EU542915 EU542916 FJ587492
ef1a	EF1a F3 EF1a R3	5'-GTGGTATCACCATTGACATTGC-3' 5'-AGAGCCTTGGGGTTGTCTTC-3'	P A C	FJ587494 FJ587495 FJ587496
gapdh	GAPDH-F1 GAPDH-R1	Pala *et al.*, 2008b	P A C	EU542913 EU542914 FJ587493

2.6. Sequence analysis and genome expression

In southern hybrid samples (PA and PAA), the presence of single genome copies (exclusively A or P) or of both genomes (identified in ambiguous positions) was determined through sequence comparison and based on the identified polymorphic sites between the two genomes. In northern populations, the absence of "nuclear non-hybrid" specimens of AA genotype impaired the identification of A-specific polymorphisms from "pure genome samples". Thus, the detection of genome specific ambiguities was based on the comparison of *S. carolitertii* sequences (C genome) with diploid and triploid hybrids, and the assessment of conserved polymorphisms at diagnostic positions. An average of two (PA) and seven (PAA) samples of *S. alburnoides* from the Tejo Basin were analysed per gene/ per organ. Among northern samples, an average number of two diploid

(CA), six triploid (CAA and CCA) and three tetraploid (CCAA and C-A-) samples of *S. alburnoides* was analysed per gene/ per organ. Genome control sequences, representing the P, A and C genomes and obtained, respectively from *S. pyrenaicus*, "nuclear non-hybrid" *S. alburnoides and S. carolitertii* were also analysed for each gene/organ combination. Two to four forward and reverse sequences of each gene were obtained per individual/ per organ, using independently synthesized cDNA samples as templates. PCR amplification for replicate reactions was also performed independently. Alignments were performed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.). According to the presence of sequence differences at informative positions, the expression of the two or only one genome type was determined for the obtained sequences.

2.7. Selection tests

To test for positive selective drives in hybrid genome interaction we started by assessing whether genome specific sequence polymorphism could result in non-synonymous amino acid substitution. The predicted amino acid sequences of P, C and A genome specific products of all six genes were aligned using ClustalW (Larkin *et al.*, 2007). Genes for which no genome-specific difference was observed at the protein level were excluded from further analysis. For the remaining genes, aligned sequences were used to construct a phylogenetic tree using the maximum-likelihood method in the DNAML program of the PHYLIP3.6 package (Felsenstein, 1989). Both the alignment and phylogenetic reconstruction were used as basis for examination of possible evolutionary models using the PAML4 package (Yang, 2007). Four site models were analysed: the M1a (Nearly Neutral), the M2a (Positive selection), the M7 (beta) and M8 (beta&ω). Results under the assumptions of each model were compared using the test statistic -2ΔlnL (ΔlnL=the difference in log likelihoods of the 2 models under comparison). Comparisons were performed between model pairs M1a and M2a and M7 and M8. The statistical significance of the test was assessed by comparison of the obtained values with a χ^2 distribution with degrees of freedom (df) equal to the difference in number of parameters between models. In parallel, site selection detection tests, based on the codon-based maximum likelihood methods FEL and REL (Pond & Frost, 2005a), were performed using the HyPhy package as implemented in the Datamonkey web server (Pond & Frost, 2005b). Nucleotide substitution models estimated with basis on the initial data set, through a maximum likelihood method were subsequently used in the site selection likelihood tests. Additionally, the selective pressures acting on different lineages were assessed using the GA-branch method (Pond & Frost, 2005c) as implemented in the Hyphy package. The method is based on the use of a genetic algorithm to infer phylogenetic relationships between lineages based on the dN/dS ratios, fitting the best model for each lineage. This method, as the other ones applied for the detection of selective pressures, shows

limitations in the analysis of hybrid samples, so an approximation was made considering individual genomes (A, P and C) as lineages. In the case of P and C genomes, although present in hybrid individuals, their contribution to the complex is made mainly through the two bisexual species with no overlapping distribution, *S. pyrenaicus* and *S. carolitertii,* so any differential selective pressures should be considered independently in the two species. Moreover, as there is no evidence of recombination between the different genomes, we considered that the A genome represented an independent genetic entity or more specifically a distinctively evolving lineage. Comparisons between the evolutionary topologies obtained from datasets containing only lineages from the *S. alburnoides* complex (P, C and A genomes) and similar ones, including gene orthologs representing an additional cyprinid lineage – the zebrafish *D. rerio*- were performed for all genes analysed with the GA-branch method**.**

3. Results

3.1. Genotyping of individuals and organs

The analysis of microsatellite variation, based on the genome-specific allele scoring information obtained for southern (Crespo-Lopéz *et al.*, 2007) and northern (Pala & Coelho, 2005; Cunha *et al.*, 2008) populations of the *S. alburnoides* complex revealed that globally the three sample sets were composed of a reasonably good representation of the forms that constitute the complex in each respective location (Table II). Diploid *S. alburnoides* from the southern Sorraia River were genotyped as PA (hybrids) and AA (nuclear "non-hybrid"). All triploids were genotyped as PAA from fin clips, and identified by the presence of three alleles for at least one locus. No triploid PPA individuals or tetraploids were found in this sample. *S. alburnoides* hybrids from the Ceira River were genotyped as CA (2), CAA (2), and CCA (1). The northern sample of the Paiva River was composed of individuals of all ploidies (diploid, triploid and tetraploid) and of representative genomic constitutions (namely the two forms of triploids- CAA and CCA- were represented in the sample). Genotypes obtained from different organs used in the analysis of gene expression patterns showed full correspondence to the ones resulting from the fin clips of the same individual, thus confirming the presence of the expected genome copies in all samples.

Table II. Summary of amplification results for the three microsatellite loci (LCO3, LCO4, LCO5) in the three River basins (Tejo, Mondego and Douro): Species (*S. alburnoides* and *S. pyrenaicus*); Sex; P (ploidy); n (number of the individual); alleles identified for each locus; and (G) genotype. An additional locus (LCO1) was used in samples for which it was not possible to obtain complete genotypes with the three loci (*).

Species	Sex	P	n	LCO3	LCO4	LCO5	G
Sorraia River - Tejo							
S. pyr	M	2n	T1	$243^{P}243^{P}$	$238^{P}238^{P}$	$137^{P}143^{P}$	PP
S. pyr	F	2n	T2	$243^{P}243^{P}$	$240^{P}242^{P}$	$137^{P}137^{P}$	PP
S. alb	M	2n	T1	$247^{A}247^{A}$	$268^{A}270^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	T2	$247^{A}247^{A}$	$270^{A}302^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	T3	$247^{A}247^{A}$	$286^{A}308^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	T4	$247^{A}247^{A}$	$278^{A}280^{A}$	$131^{A}131^{A}$	AA
S. alb	M	2n	T5	$239^{A}243^{P}$	$238^{P}268^{A}$	$131^{A}137^{P}$	PA
S. alb	M	2n	T6	$243^{P}247^{A}$	$238^{P}284^{A}$	$131^{A}143^{P}$	PA
S. alb	F	2n	T7	$243^{P}247^{A}$	$240^{P}290^{A}$	$131^{A}137^{P}$	PA
S. alb	F	3n	T8	$243^{P}247^{A}$	$238^{P}264^{A}276^{A}$	$131^{A}141^{P}$	PAA
S. alb	F	3n	T9	$243^{P}247^{A}$	$238^{P}282^{A}292^{A}$	$131^{A}141^{P}$	PAA
S. alb	F	3n	T10	$239^{A}243^{P}247^{A}$	$272^{P}296^{A}296^{A}$	$131^{A}131^{A}143^{P}$	PAA
S. alb	F	3n	T11	$243^{P}247^{A}249^{A}$	$268^{P}276^{A}276^{A}$	$131^{A}131^{A}143^{P}$	PAA
S. alb	F	3n	T12	$243^{P}247^{A}$	$238^{P}264^{A}276^{A}$	$131^{A}141^{P}$	PAA
S. alb	F	3n	T13	$243^{P}247^{A}$	$240^{P}270^{A/P}278^{A}$	$131^{A}137^{P}$	PAA*
S. alb	F	3n	T14	$243^{P}247^{A}$	$272^{A/P}282^{A}$	$131^{A}145^{P}$	PAA*
S. alb	F	3n	T15	$243^{P}247^{A}$	$238^{P}262^{P/A}290^{A}$	$131^{A}145^{P}$	PAA*
S. alb	F	3n	T16	$243^{P}247^{A}$	$238^{P}268^{P/A}292^{A}$	$131^{A}137^{P}$	PAA*
S. alb	F	3n	T17	$243^{P}247^{A}$	$238^{P}282^{A}302^{A}$	$131^{A}131^{A}137^{P}$	PAA
S. alb	F	3n	T18	$243^{P}247^{A}$	$238^{P}270^{A/P}282^{A}$	$131^{A}137^{P}$	PAA
S. alb	F	3n	T19	$243^{P}247^{A}$	$238^{P}282^{A}302^{A}$	$131^{A}131^{A}137^{P}$	PAA

Species	Sex	P	n	LCO3	LCO4	LCO5	G
Ceira River - Mondego							
S. car	F	2n	M1	243^{C}	$252^{C}256^{C}$	143^{C}	CC
S. car	F	2n	M2	243^{C}	$260^{C}264^{C}$	143^{C}	CC
S. alb	M	2n	M3	$241^{A}243^{C}$	$260^{C}272^{A}$	$131^{A}149^{C}$	CA
S. alb	M	2n	M4	$241^{A}243^{C}$	242^{C} 270^{A}	$131^{A}143^{C}$	CA
S. alb	M	3n	M5	$241^{A}243^{C}251^{A}$	$260^{C}270^{A}$	$131^{A}143^{C}$	CAA
S. alb	F	3n	M6	$241^{A}243^{C}$	$260^{C}270^{A}274^{A}$	$131^{A}149^{C}$	CAA
S. alb	M	3n	M7	$243^{c}251^{A}$	$260^{C}270^{A}$	$131^{A}143^{C}149^{C}$	CCA
Paiva River - Douro							
S. car	M	2n	D1	245^{C}	$238^{C}242^{C}$	145^{C}	CC
S. car	F	2n	D2	245^{C}	$242^{C}242^{C}$	145^{C}	CC
S. alb	?	3n	D3	245^{C} $251^{A}255^{A}$	$242^{C}262^{A}$	$133^{A}145^{C}$	CAA
S. alb	?	3n	D4	245^{C} $251^{A}255^{A}$	$242^{C}262^{A}264^{A}$	$133^{A}145^{C}$	CAA
S. alb	F	3n	D5	245^{C} 251^{A}	$242^{C}264^{A}$	$133^{A}151^{C}$	CAA*
S. alb	M	3n	D6	245^{C} 255^{A}	$238^{C}242^{C}262^{A}$	$133^{A}145^{C}$	CCA
S. alb	F	3n	D7	245^{C} 255^{A}	$238^{C}242^{C}264^{A}$	$133^{A}145^{C}$	CCA
S. alb	F	3n	D8	245^{C} 255^{A}	$242^{C}264^{A}$	$133^{A}151^{C}$	CCA*
S. alb	F?	4n	D9	245^{C} 255^{A}	$238^{C}242^{C}262^{A}264^{A}$	$133^{A}151^{C}$	CCAA
S. alb	?	4n	D10	245^{C} 255^{A}	$238^{C}264^{A}$	$133^{A}151^{C}$	CCAA*
S. alb	F	4n	D11	245^{C} 255^{A}	$238^{C}242^{C}262^{A}$	$133^{A}145^{C}$	C-A-
S. alb	F	4n	D12	245^{C} 255^{A}	$238^{C}262^{A}$	$133^{A}145^{C}$	C-A-

*

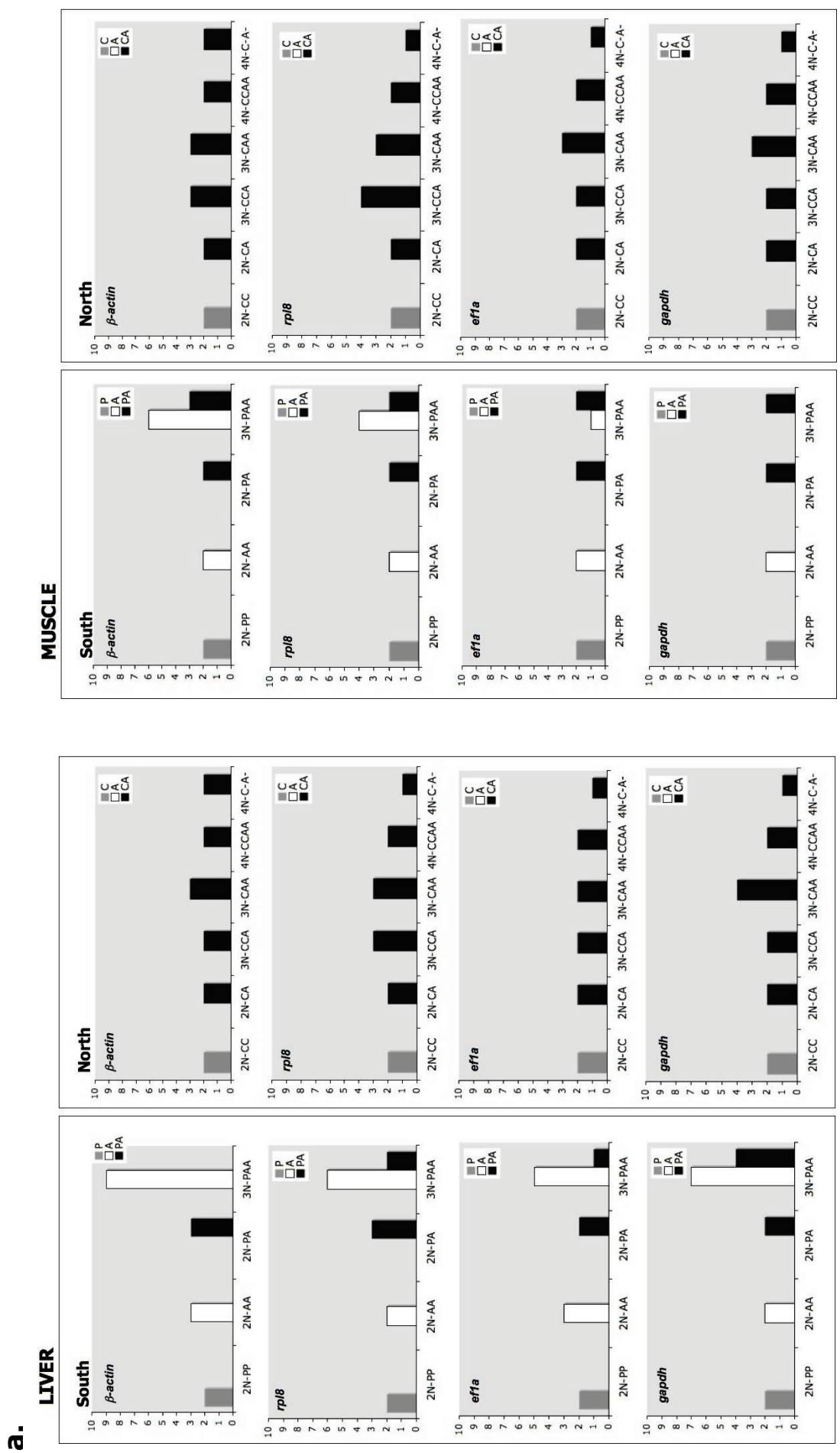
Species	Sex	P	n	LCO1
S. alb	F	3n	T13	$264^{P}364^{A}368^{A}$
S. alb	F	3n	T14	260^{P} $354^{A}374^{A}$
S. alb	F	3n	T15	$260^{P}362^{A}366^{A}$
S. alb	F	3n	T16	$256^{P}350^{A}358^{A}$
S. alb	F	3n	D5	$304^{c}334^{A}342^{A}$
S. alb	F	3n	D8	$314^{A}332^{C}352^{C}$
S. alb	F	3n	D10	$318^{A}326^{A}336^{C}352^{C}$

3.2. Gene expression patterns differ according to organ and geographical location

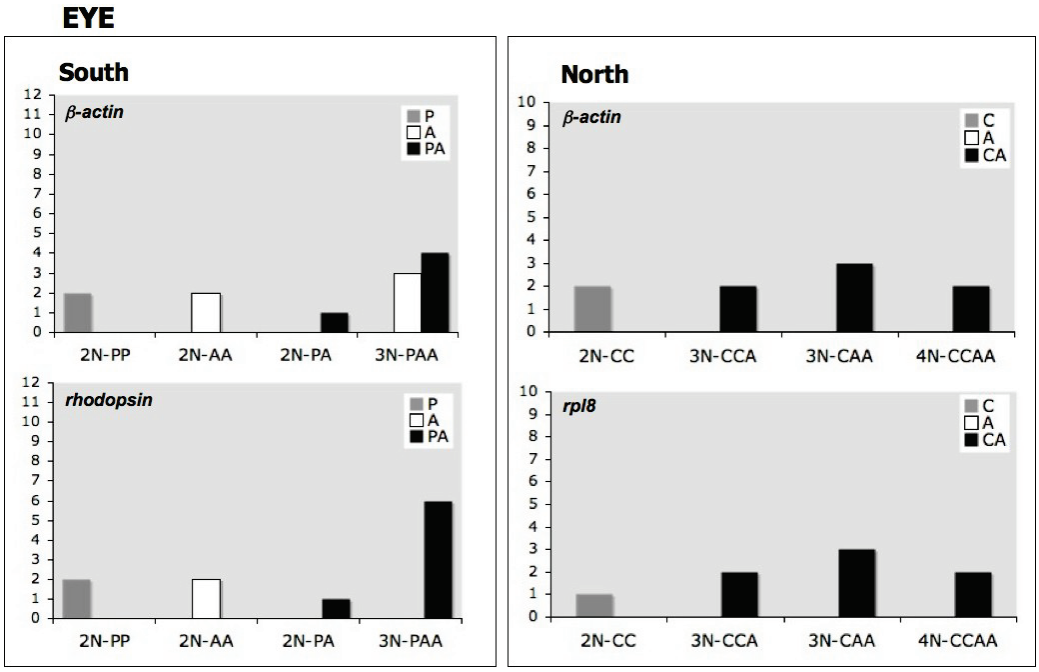
The sequence analysis of genome-specific polymorphisms in different gene and organ combinations revealed strongly contrasting results between southern and northern samples of *S. alburnoides* (Fig. 2). In the southern sample of the Tejo Basin, the four housekeeping genes (*ß-actin*, *rpl8, ef1a* and *gapdh*) showed differential expression patterns according to organ. In both muscle and liver (Fig. 2a), diploid PA individuals expressed A and P alleles, as expected. Triploids, on the other hand, exhibited a higher variety of expression profiles: for *ß-actin* and *rpl8* most muscle samples exhibited expression exclusively from the A genome, although biallelic expression was also observed; for *ef1a* two triploid samples showed biallelic expression while in the remaining one, only A genome transcripts were identified in all sequences analysed; for *gapdh,* simultaneous expression of P and A alleles was identified in all triploid samples. In liver samples all genes exhibited an overall predominance of A-genome transcripts, with a smaller number of individuals exhibiting biallelic expression. The most paradigmatic example of this was the *ß-actin* gene, of which transcripts in triploids corresponded exclusively to the A genome. In northern individuals from the Mondego and Douro Basins, biallelic expression was observed with no exceptions in all gene/organ combinations irrespective of ploidy level and genomic constitution. Thus, and as most parsimoniously expected by their hybrid genome constitution, we find qualitative evidence of expression of both genome-specific alleles for the four housekeeping genes in liver and muscle samples of CA, CCA, CAA and CCAA *S. alburnoides*. The same is true when analysing *ß-actin* and *rpl8* transcripts in eye samples of triploid (CCA and CAA) and tetraploid individuals: both C and A alleles are present. A tendency towards a preferential biallelic expression was also observed in eye samples of southern triploid (PAA) specimens: four out of seven samples showed expression of P and A alleles of the *ß-actin* gene, while *rhodopsin* transcripts resulted from biallelic expression in all samples analysed.

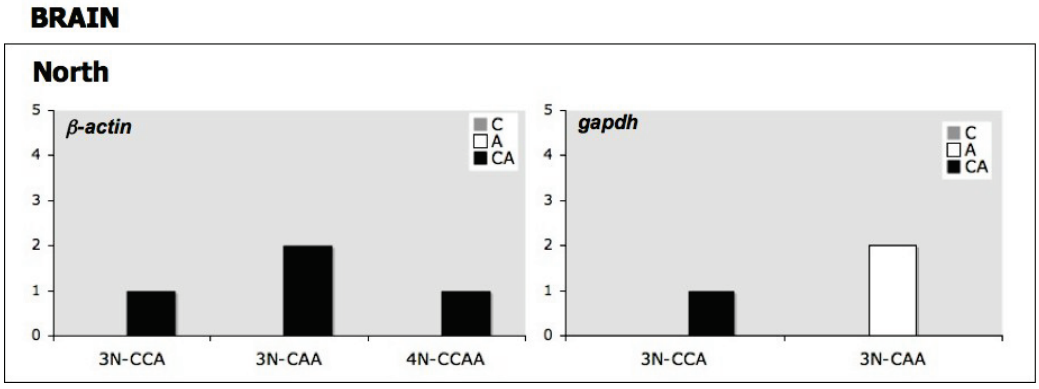
Brain samples from triploid and tetraploid individuals from Mondego and Douro showed a contrasting feature to the overall tendency of biallelic expression of C and A genomes in these basins: although C and A *ß-actin* expression was observed in all samples, the *gapdh* gene showed monoallelic, A genome-exclusive expression in CAA individuals (Fig. 2b).

In adult gonads (Fig. 2c) of southern triploid individuals of *S. albunoides*, all four housekeeping genes (*ß-actin*, *rpl8*, *ef1a* and *gapdh*) showed exclusive expression of A genome alleles. *Amh* showed biallelic expression in triploid PAA, while *dmrt1* exhibited both monoallelic (9 samples) and biallelic (3 samples) genome expression. Diploid PA exhibited a biallelic expression pattern in all genes analysed (*amh, dmrt1* and *ß-actin*). In northern samples, patterns were divergent according to gene and genomic constitution.


a.
LIVER
MUSCLE
South
North
β-actin
rpl8
ef1a
gapdh
2N-PP
2N-AA
2N-PA
3N-PAA
2N-CC
2N-CA
3N-CCA
3N-CAA
4N-CCAA
4N-C-A-
P
A
PA
C
CA

b.


EYE
South
β-actin
P
A
PA
2N-PP
2N-AA
2N-PA
3N-PAA
rhodopsin
North
β-actin
C
A
CA
2N-CC
3N-CCA
3N-CAA
4N-CCAA
rpl8


BRAIN
North
β-actin
C
A
CA
3N-CCA
3N-CAA
4N-CCAA
gapdh

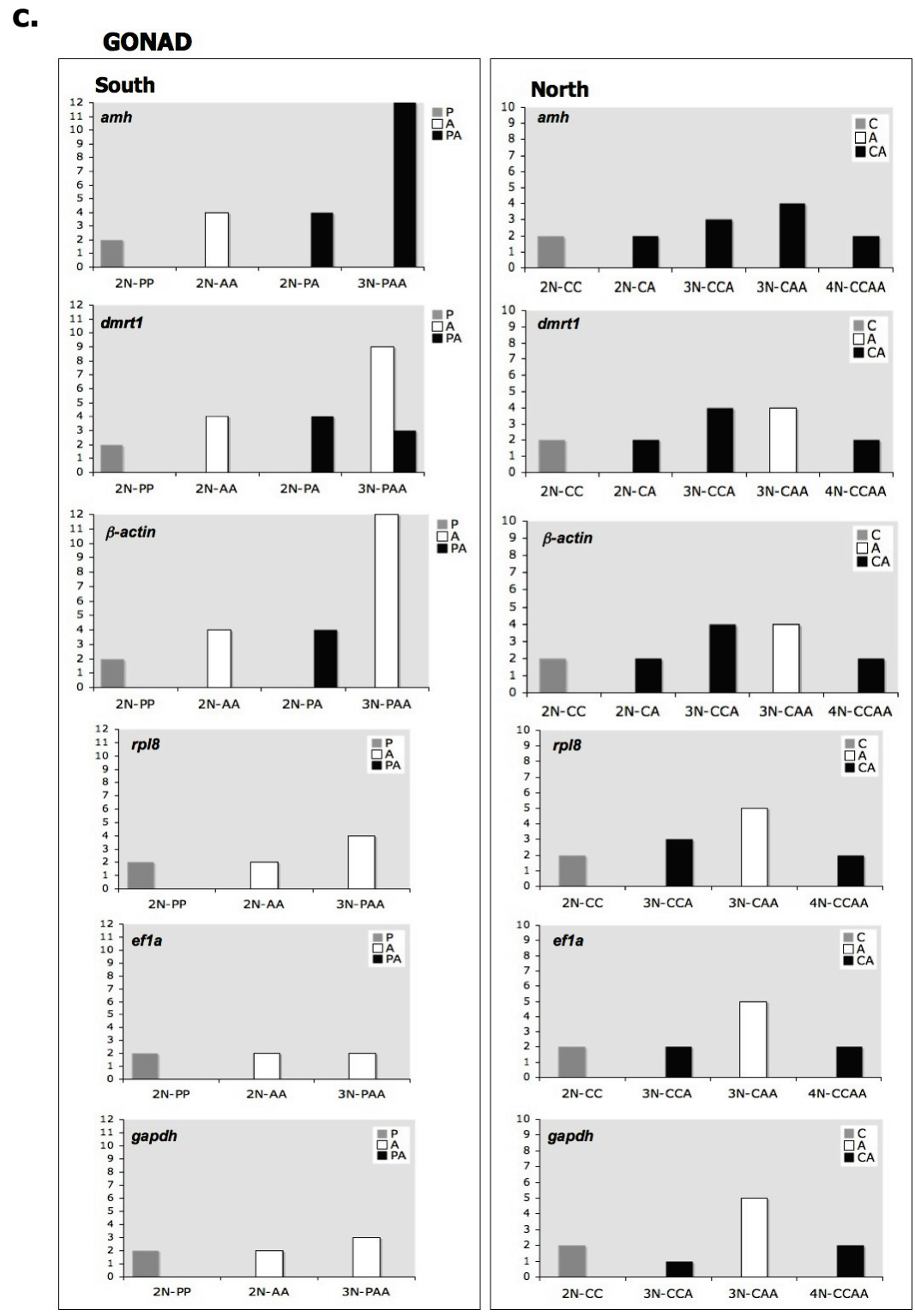


Figure 2. Frequency of allele-specific transcripts of P, C and A in different organs of individuals from southern (Tejo) and northern (Mondego and Douro) populations of the complex. (a) *ß-actin, rpl8, ef1a* and *gapdh* in muscle and liver; (b) *ß-actin, rhodopsin, rpl8 in* eye and *ß-actin* and *gapdh* in brain; (c) *amh*, *dmrt1*, *ß-actin, rpl8, ef1a* and *gapdh* in gonad. The number of samples showing expression of the A, P or both genomes is shown in the histogram, on the Y-axis. The genotype of each group of samples is indicated on the X-axis.

Amh was the only gene for which expression of both C and A alleles was observed in all diploid, triploid and tetraploid forms of *S. alburnoides*. The remaining five genes exhibited a distinctive expression profile between the two forms of triploids: while in CCA samples the observed expression pattern was biallelic (in accordance to diploid and tetraploid samples from the same locations), in CAA individuals, *dmrt1*, *ß-actin*, *rpl8*, *ef1a* and *gapdh* expression was due exclusively to the A genome.

3.3. Selection

The comparison of the predicted amino acid sequences of the seven genes analysed in the present work revealed that in many cases, the genome specific polymorphisms observed in DNA sequence corresponded to synonymous substitutions, thus causing no changes in the resulting proteins. Rhodopsin, Rpl8, Ef1a and Actin predicted proteins showed only synonymous aminoacid substitutions and a conserved amino acid sequence when comparing P, C and A genomes. Non-synonimous substitutions were detected in the Gapdh, Amh and Dmrt1 proteins within the three genomes that interact in the different populations of the *S. alburnoides* complex (Supplementary Fig. 1). Amh sequences were the only ones for which the number of non-synonymous susbtitutions was higher than the number of synonymous ones. For the other two genes, the number of non-synonymous sites was always equal or lower than the number of synonymous sites. The three genes were selected for further analysis to assess whether selective pressures could account for their distinctive expression patterns in different populations of the complex. No sites under positive selection were detected by the analysis with the PAML4 package. There were no differences in the log likelihood values of the model pairs under comparison (M1a-M2a; M7-M8) in all genes analysed. Similar results, implying the absence of positively selected sites, were obtained when applying the FEL and REL methods as implemented in the HyPhy package. In this analysis, a more extended set of six genes (*amh*, *dmrt1*, *ß-actin*, *rpl8*, *ef1a* and *gapdh*) was used to assess whether other types of selective forces, rather than positive selection could be acting in individual sites. Aminoacid positions under negative selection were detected in *ß-actin*, *ef1a* and *gapdh* at a significance level of 0.05. In all genes, rate classes showed dN<dS, thus suggesting that genes are under purifying selection. The topology obtained by the application GA branch methods suggested segregation between the P, C and, more interestingly, between “southern and northern type” A genomes. However, the scarce polymorphism between genomes and the low number of sequences used for both selection and GA branch analysis might provide only low statistical support for the obtained results. We thus considered the results as indicative, rather than statistically significant.

4. Discussion

In the present work we have addressed the question of gene expression regulation following hybridisation and ploidy rise by studying the expression patterns of seven housekeeping and tissue specific genes in individuals of different genomic composition and ploidy degrees, from different locations within the distribution range of the *S. alburnoides* complex. Our aim was to assess whether differences in the evolutionary history and in the genomes taking part in the hybridisation process could be related to specific patterns of gene expression regulation, in different populations of the complex. Our strategy was to analyse representative samples from three major Basins within the complex distribution range, for which different evolutionary scenarios have been proposed (Pala & Coelho, 2005; Crespo-López *et al.*, 2007; Cunha *et al.*, 2008). The question would be whether similar hybridisation events, but different evolutionary trajectories, could have different functional outcomes in terms of gene expression. Additionally we aimed to assess whether the functional balance attained in each population could have a role in its evolutionary prospects and in the possibilities of perpetuation of particular lineages.

4.1. Differential gene expression patterns in southern and northern populations of *S. alburnoides*

The data obtained in the present work implies a clear difference in the regulation of gene expression correlated to geographical location of individuals: in southern populations, preferential expression of A genome, and silencing of P genome alleles was observed for the majority of genes analysed. Patterns apparently vary according to gene and organ, implying plasticity, rather than directionality in the mechanisms involved in gene silencing. A completely different scenario was observed for the samples of the northern river basins: in the majority of cases both C and A genome alleles were expressed in the gene/organ combinations analysed, irrespective of ploidy level or genomic composition. Also, no differences were found between the two northern rivers regarding expression patterns, with all forms showing the same bi-allelic expression profile. However, distinctive patterns of gene expression were observed in the gonads of the two triploid forms of *S. alburnoides* (CAA and CCA) from the two northern basins. The patterns of gene expression in CAA gonads were similar to the ones observed in southern PAA samples, thus implying that they could be related to the type of reproduction and mechanism of gamete production of the different forms. PAA and CAA females of *S. alburnoides* have been shown to produce haploid and more rarely diploid gametes in which only A genome is present (Alves at al., 1998; Carmona *et al.*, 1997; Pala & Coelho, 2005). The observed expression patterns in this triploid form correlate well with the described mechanisms: *amh*, which would be mostly expressed by the somatic cell lineage, (Pala *et al.*, 2008a) could have a biallelic gene

expression pattern, while germ cells (which at this point would constitute the largest fraction in the gonad) would mainly express A genome alleles upon the process of meiotic hybridogenesis in which the C or the P genomes are discarded from gametogenesis. Following the same reasoning for the analysis of the other triploid form, only two explanations could be proposed for the difference in expression patterns observed in CCA females: either the process of gametogenesis in these females would involve no elimination of the minority genome (which in this case would be the A genome) or the differences could be related to the maturation state of the analysed gonads (if gametogenesis would still be at a very early stage, biallelic expression in the majority of cells would be a strong possibility).

Due to their low number in most populations, no direct information about the reproductive modes of CCA females was obtained until now, but it has been hypothesized that a similar mechanism of meiosis or hybridogenetic meiosis could also be taking place during their gametogenesis, as suggested for the southern PPA form (Crespo-López *et al.*, 2006). The patterns of biallelic expression observed here for all genes might imply another explanation. Either all genomes are being transmitted to the progeny, as in triploid PAA and PPA males of southern populations (Alves *et al.*, 1999) or, if any is eliminated from the process of gamete formation it is not the minority genome. The fertility and the exact contribution of this form to the complex is yet to be determined, and crosses involving tetraploids and triploid CCA individuals have resulted in the non-development of the progeny (Cunha *et al.*, 2008). The hypothesis of differences in the maturation stage of the ovaries of CAA and CCA females is less probable, as morphologically they did not present apparent relevant differences, but it cannot be totally excluded. A more intensive analysis of various stages of ovarian maturation with more samples of each triploid form would be necessary for a more precise assessment. Only then we could have a better indication of the cell types that would be contributing to overall expression in relation to the reported mechanisms of gamete production and obtain a better perspective of the exact meaning of the observed patterns.

4.2. Gene expression and evolution of the different populations of the complex

The interpretation of these results is far from trivial, and we should start by analysing the patterns of southern and northern basins separately. In southern populations of the Tejo River Basin, the instability of uneven genome numbers is apparently overcome by genome-specific allele silencing, similarly to what has been observed in other organisms following polyploidisation (Adams *et al.*, 2003, 2004). Moreover, the observed allele silencing could be correlated to gene dosage compensation in triploids of *S. alburnoides* in this river basin (Pala *et al.*, 2008b). In plants it has been proposed that adapted polyploids would avoid extinction by reducing gene redundancy and thus

undergoing an evolutionary pathway leading to functional diploidisation (Paterson *et al.*, 2004; Wang *et al.*, 2005). The need for correct gene balance in a polyploid context as a way to tolerate the presence of extra genome copies and to maintain stable aneuploids has been further demonstrated in plants (Bircher & Veita, 2007) and accounted for as a strong possibility in vertebrates (Mable, 2007). In addition to the referred functional outcomes, effective epigenetic silencing of genes in polyploid genomes might have strong effects on their evolutionary potential (Adams & Wendel, 2005), so regulation of gene expression as a response to odd ploidy might be one of the factors contributing to the reported adaptability and evolutionary success of populations of the complex in its southern distribution (Crespo-López *et al.*, 2007). Moreover, epigenetic regulation phenomena might increase diversity, plasticity and heterosis, actively contributing to the adaptative potential of polyploid lineages (Paterson, 2005; Salmon *et al.*, 2005). However, if genome specific allele silencing and variable gene expression patterns in different organs might be desirable features in an evolutionary point of view, why would northern hybrids show strict biallelic expression patterns? Our first hypothesis would be that differential selective pressures acting on the different genomes according to geographical location could have modulated gene expression patterns accordingly. The results of the selection tests, however, do not seem to point directly towards this direction. Although northern and southern genome lineages were separated by branch forming methods, the same type of selective force was apparently acting over all genes, in all genome types (A, P and C) from the three locations. The genes analysed here are apparently all under purifying selection, which would seem in accordance with one of the possible evolutionary fates of gene duplicates, but due to the low inter-genomic variability results could only be considered as indicative, and a correlation between ploidy rise and the selective forces acting on genomes could not be directly assumed. The genes analysed are either ubiquitously expressed in accordance to their “housekeeping” functions or, in the case of the gonad-specific genes (*amh* and *dmrt1*), participate in strictly regulated cascades, in which correct interaction with partners is mandatory (Josso & Clemente, 2003; Smith *et al.*, 1999), thus it is expected that they are subjected to such negative selective forces, regardless of polyploid or hybrid context.

Nevertheless, differences at the protein level do occur between C and A genomes in northern populations. In diversity-deprived populations, such as the ones of the Mondego River basin (Pala & Coelho, 2005) the biallelic expression of genes could in fact bring about some advantages. Gene duplication has been proposed as an important factor catalyzing gene function diversification, by providing relief from a background of purifying selection (Chain *et al.*, 2008). In fact, pathways of polyploid lineage evolution often involve shorter-term acting mechanisms such as epigenetic gene

silencing (Adams & Wendel, 2005) and slower purifying selection effects on duplicates, following gene copy rise (Wendel, 2000).

The northern Mondego River populations have been reported to be in a worrying situation in terms of evolutionary potential, with less capacity of diversity maintenance and speciation than other populations of the complex (Pala & Coelho, 2005). While southern populations of *S. alburnoides* have been regarded as examples of viability among polyploid hybrid taxa, the Mondego populations have been considered as more proximate to an evolutionary "dead-end" (Pala & Coelho, 2005). Conversely, other populations within the northern distribution of the complex seem to have undergone different pathways to promote persistence and evolutionary potential, which are apparently not dependant on specific expression patterns. Tetraploids are relatively rare among the complex, but in the Douro River basin, the occurrence of populations composed almost exclusively by this form has been reported as a possible stepping-stone to the establishment of an independently reproducing, bisexual lineage (Cunha *et al.*, 2008). These tetraploid *S. alburnoides* are reproductively isolated from other forms of the complex and their pair genome number and symmetrical genotypic constitution allows normal meiosis and recombination, so it is possible that an ongoing speciation process is being observed in the Douro River basin (Cunha *et al.*, 2008).

Could this indicate that different lineages within the same hybrid complex could have adopted distinct functional strategies to overcome the molecular instability brought upon by hybridisation? Would these pathways then directly intermingle, result from or even determine their particular evolutionary fates? In the Tejo River basin, populations of this hybrid complex have been shown to be fully adapted and viable (Crespo-López *et al.*, 2007). The most successful form of *S. alburnoides* in these locations is the triploid, precisely the one in which we have observed the genome-specific allele silencing phenomena. Thus, it is possible that in these populations, the perpetuation of gene copies in a polyploid context would have been attained through balanced expression and that this process could be actively contributing to the evolutionary success of southern *S. alburnoides*. In other populations of the complex, this balance could still be attained, as simultaneous expression of heterologous alleles does not imply that all three or four gene copies are being expressed, only that there is no preferential silencing of alleles of one genome. In populations with few opportunities of maintaining diversity and in an impoverished situation in terms of evolutionary potential, such as the ones of the Mondego River basin (Pala & Coelho, 2005), this pattern could be advantageous, by maintaining genes in a "permanent" state of heterozygosis. But the complex seems to have undergone other pathways of polyploid lineage perpetuation. If in southern populations the "diploidisation" necessity for the maintenance of gene duplicates was accomplished by epigenetic regulation of gene expression, in some northern populations of the Douro River basin, other

pathways are apparently being followed towards lineage perpetuation and speciation. In these populations, the molecular instability of odd genome number seems to have been overcome by the constitution of symmetrical tetraploid lineages, with a closer functional resemblance to a diploid. The return to an even ploidy, be it effectively or only functionally, is an important process for genome stabilization and classically presented as an evolutionary "fuel" in polyploid evolution of various taxa (Henry *et al.*, 2005). According to this perspective, one could propose that the triploid form of *S. alburnoides* would act as an intermediate evolutionary step in the classically considered pathway of redundant gene perpetuation and diversification (Ohno, 1970; Meyer & Schartl, 1999).

4.3. Hybridisation processes and the involvement of different genomes

What could be in the origin of such marked difference? It is not clear, as the different populations of the complex have been originated by similar events and maintained by analogous mechanisms of genetic exchange. One possibility could be the difference in the origin and the timing of constitution of the three lineages of the complex analysed here. Regulation of gene expression can be variable in ancient and newly formed polyploids (Wang *et al.*, 2004; Adams & Wendel, 2005) and the disparity in the timing of the onset of hybridisation events in the different lineages of the *S. alburnoides* complex could be a factor to be taken into account. The proximity, in terms of evolutionary time, of the complex's dispersal in these three river basins (raging from less than 0.7MY to 0.01 MY) (Sousa-Santos *et al.*, 2007) and the involvement of the same genome combinations in the onset of the hybridisation process in all locations, if we consider independent origins (Cunha *et al.*, 2004), makes this hypothesis improbable. However, differences emerge when considering the composition and the genomes involved in the progression of the complex in each of the three basins. The C genome, introduced through the bisexual species *S. carolitertii* in northern populations was a later addition to the complex, after initial hybridisation events involving only *S. pyrenaicus* (Alves *et al.*, 1997). Genomic stress induced by two distinct genomes being brought together in the same cell nucleus often leads to different epigenetic modifications of homeologous, but the effects might vary according to the intrinsic characteristics of each genome involved in the hybridisation process. Among other processes, silencing or downregulation can result from the interaction between divergent regulatory hierarchies (Riddle & Birchler, 2003) and differential capacity of interaction with other proteins or complexes (Comai, 2000; Adams & Wendel, 2004). At a more restricted level, phenomena such as nuclear dominance, in which only one parental genome contributes and "dominates" the formation of nucleoli, have been associated with hybridisation events (Chen *et al.*, 1998; Flowers & Burton, 2006). Both the P and C genome apparently have a good functional "affinity" with the A genome, but it can be possible that regulatory elements that have to interact

with both heterologous genomes in hybrids, respond differentially or have distinct interaction capacity.

Another striking difference between the northern and southern populations of the complex, which has been pointed out as a key factor in the maintenance of its diversity and contributing actively to the evolutionary success of *S. alburnoides* is the presence of "nuclear non-hybrid" males of AA genotype (Alves *et al.*, 2002; Crespo-López *et al.*, 2006, 2007). In southern populations, early in the formation of the complex, A oocytes produced from triploid females (formed by mating of PA hybrid females and the paternal ancestor of the complex) would have been fertilized by gametes of the paternal ancestor, giving rise to the "nuclear non-hybrid" lineage. These events might have initiated the cycle of "non-hybrid" lineage perpetuation, through diploid and triploid females and the establishment of this lineage as representative of the ancestor genome when the paternal species was lost from those locations (Alves *et al.*, 2002). Apparently, such pathway did not take place in northern lineages and we can only speculate about the reasons why this could have happened. It could be possible that when a relevant number of A oocytes would have been produced by triploids, the parental species had been already lost from these locations, thus impeaching syngamy of A gametes and lineage initiation. Alternatively, the lineage could have been established, as in southern populations, but subsequently been wiped out from these locations or overcome by the other forms of the complex. Thus, the regulatory dynamics of A genome alleles in southern populations could be related to the establishment of the nuclear "non-hybrid" lineage and the introduction of A alleles originating from this form, that somehow might have evolved differentially comparatively to the paternal ancestor. A different type of dynamics could have been established in northern populations in which this process of A allele recombination and shifting is not present. Whether or not this effect is emphasised by differences in the regulatory behaviour of P and C genomes is yet to be determined. It also remains to be established whether or not dosage compensation effects are also present in northern populations, despite the qualitative differences observed in the expression of the different genomes according to geographical location.

By reporting these findings, we hope to have further contributed to the idea that there is not a unique and direct solution to the question of how a hybrid polyploid faces the challenges of evolution, diversity maintenance and capacity of functionally overcoming the instability of a heterologous genome context.

Although a huge diversity of processes and evolutionary "approaches" brought upon hybridisation and ploidy rise have been described in plants, such mechanisms are still not understood and scarcely studied in polyploid animals. By showing that differential expression patterns can occur within the same polyploid complex and that this heterogeneity could be related to the evolutionary history of

populations, we expect to open the door to further investigation of how vertebrate polyploid genomes are regulated and how the different adopted strategies can influence their odds on the race for adaptability, evolutionary success and speciation.

References

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U.S.A. 100, 4649-4654.

Adams, K.L., Wendel, J.F., 2004. Exploring the genomic mysteries of polyploidy in cotton. Biol. J. Linn. Soc. 82, 573-582.

Adams, K.L., Percifield, R., Wendel, J.F., 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168, 2217-2226.

Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. Curr Opin Plant Biol. 8, 135-141.

Alleman M, Doctor J., 2000. Genomic imprinting in plants: observations and evolutionary implications. Plant Mol Biol. 43,147-161.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J. Dowling, T.E., 1997. Maternal ancestry of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae) as determined by analysis of cytochrome *b* sequences. Evolution 51, 1584-1592.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 1998. Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. Mol. Biol. Evol. 15, 1233-1242.

Alves, M.J., Coelho, M.M., Próspero, M.I., Collares-Pereira, M.J., 1999. Production of fertile unreduced sperm by hybrids males of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): an alternative route to genome tetraploidization in unisexuals. Genetics 151, 277-283.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111, 375-385.

Alves, M.J., Collares-Pereira, M.J., Dowling, T.E., Coelho, M.M., 2002. The genetics of maintenance of an all-male lineage in the *Squalius alburnoides* complex. J Fish Biol 60, 649–662

Alves, M. J., Gromicho, M., Collares-Pereira, M. J., Crespo- López, E., Coelho, M. M., 2004. Simultaneous production of triploid and haploid eggs by triploid *Squalius alburnoides* (Teleostei: Cyprinidae). J. Exp. Zool. 301A, 552–558.

Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H., Birchler, J.A., 2005. Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics 169, 389-397.

Birchler, J.A., Veitia, R.A., 2007. The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19, 395–402.

Carmona, J.A., Sanjur, O.I., Doadrio, I., Machordom, A. Vrijenhoek, R.C., 1997. Hybridogenetic reproduction and maternal ancestry of polyploid Iberian fish: the *Tropidophoxinellus alburnoides* complex. Genetics 146, 983-993.

Chain, F.J., Ilieva, D., Evans, B.J., 2008. Duplicate gene evolution and expression in the wake of vertebrate allopolyploidisation. BMC Evol Biol. 8, 43.

Chen, Z.J., Comai, L., Pikaard, C.S., 1998. Gene dosage and stochastic effects determine the severity and direction of uniparental rRNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. Proc Natl Acad Sci USA 95, 14891–14896.

Comai, L., 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol 43, 387-399.

Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., Byers, B., 2000. Phenotypic instability and rapid gene silencing in newly formed *arabidopsis* allotetraploids. Plant Cell 12, 1551-1568.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6, 836-846.

Crespo-López, M.E., Duarte, T., Dowling, T., Coelho, M.M., 2006. Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. Zoology (Jena) 109, 277-286.

Crespo-López, M.E., Pala, I., Duarte, T., Dowling, T.E., Coelho, M.M., 2007. Genetic structure of the diploid–polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. J. Fish Biology 71, 423-436.

Cunha, C., Coelho, M. M., Carmona, J. A., Doadrio, I. 2004 Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridisation events. Mol. Ecol. 13, 2807–2817.

Cunha, C., Doadrio, I., Coelho, M.M., 2008. Speciation towards tetraploidization after intermediate processes of non-sexual reproduction. Philos Trans R Soc Lond B Biol Sci. 363, 2921-2929.

Dawley, R.M., Goddard, K.A., 1988. Diploid-triploid mosaics among unisexual hybrids of the minnows *Phoxinus neogaeus*. Evolution 42, 649-659.

Dowling, T. E., Secor, C. L., 1997. The role of hybridisation and introgression in the diversification of animals. Annu. Rev. Ecol. Syst. 28, 593–620.

Felsenstein, J., 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

Fitzgibbon, J., Hope, A., Slobodyanyuk, S.J., Bellingham, J., Bowmaker, J.K., Hunt, D.M., 1995. The rhodopsin-encoding gene of bony fish lacks introns. Gene 164, 273-277.

Flowers, J.M., Burton, R.S., 2006. Ribosomal RNA gene silencing in interpopulation hybrids of *Tigriopus californicus*: nucleolar dominance in the absence of intergenic spacer subrepeats. Genetics 173, 1479-1486.

Guo, M., Davis, D., Birchler, J.A., 1996. Dosage effects on gene expression in a maize ploidy series. Genetics 142, 1349-1355.

Henry, I.M., Dilkes, B.P., Young, K., Watson, B., Wu, H., Comai, L., 2005. Aneuploidy and genetic variation in the *Arabidopsis thaliana* triploid response. Genetics 170, 1979-1988.

Hotz, H., Semlitsch, R.D., Gutmann, E., Guex, G.D., Beerli, P., 1999. Spontaneous heterosis in larval life-history traits of hemiclonal frog hybrids. Proc Natl Acad Sci USA 96, 2171-2176.

Josso, N., Clemente, N., 2003. Transduction pathway of anti-Mullerian hormone, a sex-specific member of the TGF-beta family. Trends Endocrinol Metab 14, 91–97.

Lampert, K.P., Lamatsch, D.K., Fischer, P., Epplen, J.T., Nanda, I., Schmid, M., Schartl, M., 2007. Automictic reproduction in interspecific hybrids of poeciliid fish. Curr Biol. 17, 1948-1953.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G., 2007.ClustalW and ClustalX version 2. Bioinformatics 2007 23, 2947-2948.

Mable, B.K., 2007. Sex in the postgenomic era. Trends Ecol Evol. 22, 559-561.

Meyer, A., Schartl, M., 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr. Opin. Cell. Biol. 11, 699-704.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16, 1215.

Ohno, S., 1970. Evolution by Gene Duplication. George Allen and Unwin, London.

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401-437.

Pala, I., Coelho, M. M., 2005. Contrasting views over a hybrid complex: between speciation and evolutionary "dead-end". Gene 347, 283–294.

Pala, I., Klüver, N., Thorsteinsdottir, S., Schartl, M., Coelho, M.M., 2008a. Expression pattern of anti-Müllerian hormone (*amh*) in the hybrid fish complex of *Squalius alburnoides*. Gene 410, 249-258.

Pala, I., Coelho, M.M, Schartl, M., 2008b. Dosage compensation by gene copy silencing in a triploid hybrid fish. Curr Biol. 18, 1344-1348.

Paterson, A.H., Bowers, J.E., Chapman, B.A., 2004. Ancient polyploidisation predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl Acad. Sci. USA 101, 9903–9908.

Paterson, A.H., 2005. Polyploidy, evolutionary opportunity, and crop adaptation. Genetica 123, 191–196.

Pignatta, D., Dilkes, B., Wroblewski, T., Michelmore, R.W., Comai, L., 2008. Transgene-Induced Gene Silencing Is Not Affected by a Change in Ploidy Level. PLoS ONE 3(8): e3061.

Pond, S.L., Frost, S.D., 2005a. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22, 1208-1222.

Pond, S.L., Frost, S.D., 2005b. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 21, 2531-2533.

Pond, S.L., Frost, S.D., 2005c. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol. 22, 478-485.

Riddle, N.C., Birchler, J.A., 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Genet 19, 597-600.

Salmon, A., Ainouche, M.L., Wendel, J.F., 2005. Genetic and epigenetic consequences of recent hybridisation and polyploidy in *Spartina* (*Poaceae*). Mol. Ecol. 14, 1163–1175.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., Levy, A.A., 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridisation and allopolyploidy in wheat. Plant Cell 13, 1749-1759.

Smith, C.A., McClive, P. J., Western, P.S., Reed, K.J., Sinclair, A. H., 1999. Conservation of a sex-determining gene. Nature 402, 601– 602.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V., 2007. Reading the history of a hybrid fish complex from its molecular record. Mol Phylogenet Evol. 45, 981-996.

Stöck, M., Lamatsch, D.K., Steinlein, C., Epplen, J.T., Grosse, W.R., Hock, R., Klapperstück, T., Lampert, K.P., Scheer, U., Schmid, M., Schartl, M., 2002. A bisexually reproducing all-triploid vertebrate. Nat Genet. 30, 325-328.

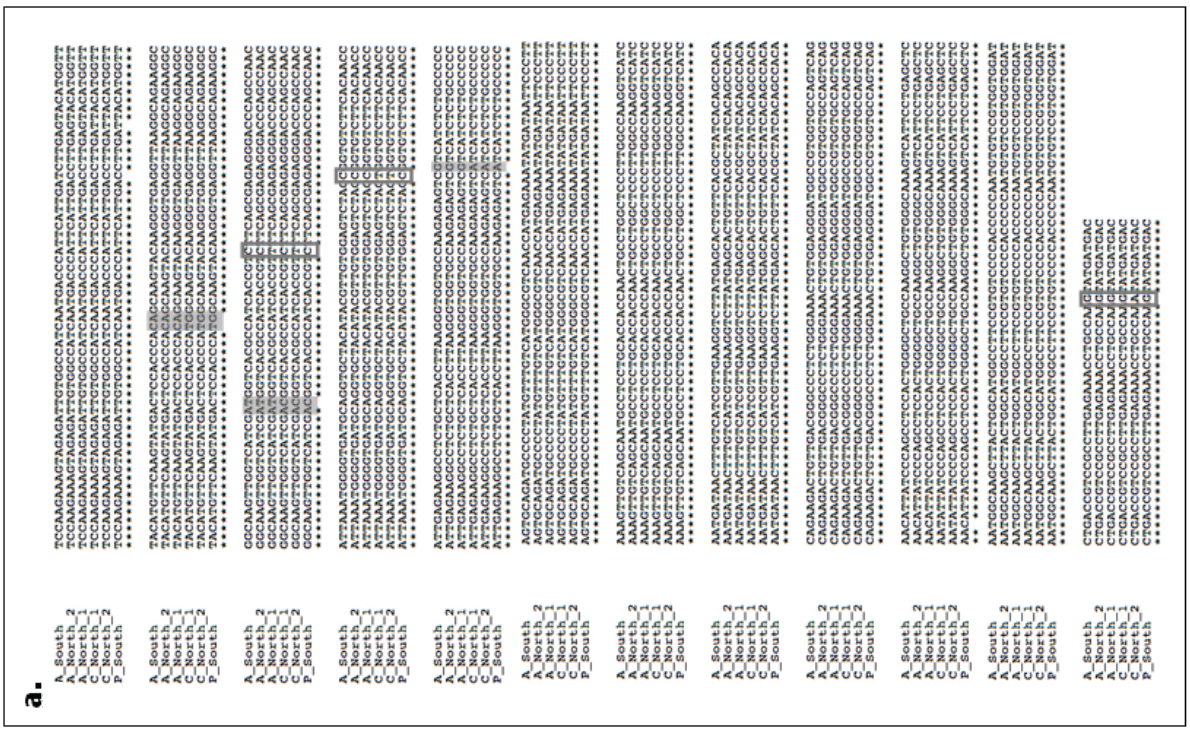
Wang, J., Tian, L., Madlung, A., Lee, H.S., Chen, M., Lee, J.J., Watson, B., Kagochi, T.,Comai, L., Chen, Z.J., 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics 167, 1961-1973.

Wang, X., Shi, X., Hao, B., Ge, S., Luo, J., 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidisation. New Phytol. 165, 937–946.

Wendel, J.F., 2000. Genome evolution in polyploids. Plant Mol Biol. 42, 225–249.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24,1586-1591.

CHAPTER 5 | Supplementary data


a.
b.
Gapdh

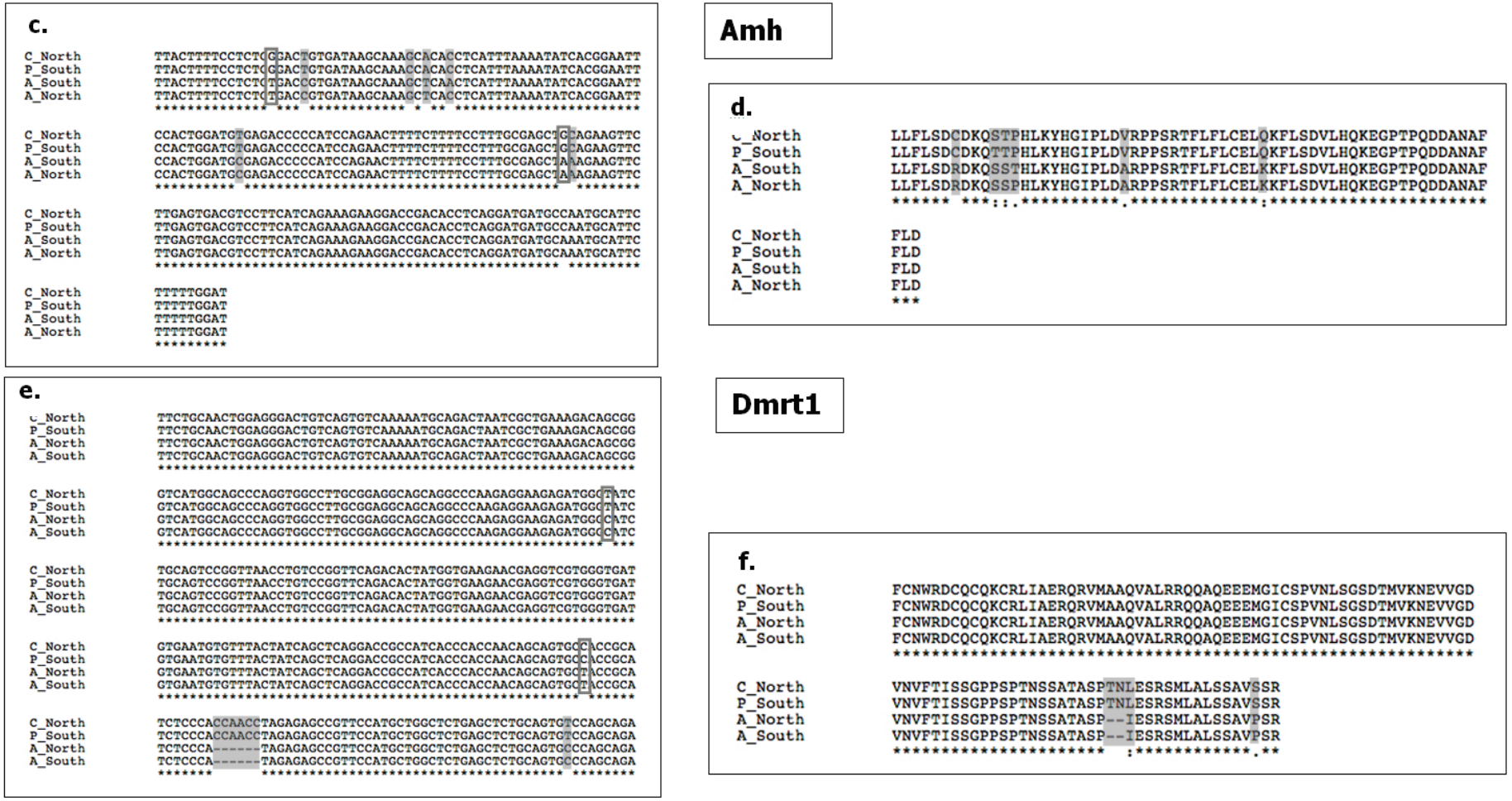


Figure S1. Polymorphic sites in the coding sequences of the *gapdh*, *amh* and *dmrt1* genes (a., c., e.), between the P, C and A genomes. Position of non-synonymous substitutions and the corresponding aminoacid change at the protein level (b., d., f.) are highlighted in light grey (full)

CHAPTER 6 | Discussion

6. DISCUSSION

The aim of this chapter is to provide an integrative discussion of the results obtained in the four manuscripts that compose Chapters 2 to 5. Although partial discussions have already been presented regarding the findings of each individual work, in this chapter they are integrated in a more global overview and debated in light of their broader implications.

6.1. Creation of a molecular basis for the study of gene expression in *S. alburnoides*

One of the major focuses of this thesis was to clarify the impact of hybridisation and polyploidisation processes on hybrid genes and genomes, using the *S. alburnoides* fish complex as a case study to elucidate some of these features. However, for such analysis to be performed, a starting point of genetic information would have to be established. Three main strategies are possible to gather such information: the search for traces of structural rearrangements brought upon by the merging of two different chromosomal systems that have from then on to "work" together (Pontes *et al.*, 2004; Lim *et al.*, 2008); the wide-scale analysis of gene behavior throughout whole genomes (Udall *et al.*, 2006; Stupar *et al.*, 2007) which is more common in organisms for which there is reasonable amount of sequence data; and the analysis of the behavior of individual genes, be it in specific pathways or in more widespread expression in the whole organism (Adams *et al.*, 2003, 2004; Guo *et al.*, 2004; Auger *et al.*, 2005).

Attempts have been made to elucidate wider scale genomic rearrangements following the hybridisation process in *S. alburnoides* (Gromicho, 2006), but until now no systematic search for gene candidates had ever been performed. In fact, in the case of a vertebrate system like *S. alburnoides*, in which the molecular information basis is presently restricted to mitochondrial genes (Alves *et al.*, 1997; Cunha *et al.*, 2004), non-coding sequences with widespread distribution in the genome (Pala & Coelho, 2005; Crespo-López *et al.*, 2006) and the nuclear *β-actin* gene (Robalo *et al.*, 2006), a good option would be to look for gene candidates, using known genomes with reasonable evolutionary proximity as basis for the search (such as the one of the closely related cyprinid zebrafish). However, these genes could not be picked up randomly and would have to present features of interest for the global analysis. In the case of sex determination, a final set of six genes was gathered (*amh*, *dmrt1*, *vasa*, *wt1*, *dax1* and *figla*), in which all isolates respected the characteristics of gonad expression in the adult and the presence of domains essential for their described functionality. Regarding the investigation of the global patterns of gene expression, the major concern was to obtain gene products either with tissue specific or widespread expression in

which the contribution of the different genomes (for instance P and A in southern populations of the complex) could be distinguished at sequence level. A final set of four housekeeping (*ß-actin*, *rpl8*, *gapdh*, *ef1a*), and four tissue specific (*rhodopsin*, *amh*, *dmrt1* and *vasa*) genes was put together, with all genes exhibiting diagnostic polymorphisms for each of the two heteromorphic genomes represented in the hybrids, while other isolates such as partial coding sequences of the *5S* and *28S* ribosomal genes (Appendix 1) were not studied due to the absence of such sequence difference. The possibility of an innovative gene expression analysis in the context of vertebrate polyploids was therefore opened by the outstanding characteristics of *S. alburnoides* and the informative potential of the genes isolated here for the first time. This has been the most extensive gene search effort ever performed in this fish complex, and the potential applications of some of these isolates are not only restricted to the initial objectives of this thesis. At the genomic level, genes such as *amh* and *ef1a* have proven to have a good informative potential for phylogenetic analysis *in S. alburnoides*. Additionally, intronic and exonic regions of seven of these genes are also being applied by other researchers in the group as molecular markers in the analysis of different Iberian fish genera.

6.2. Genetic basis of sex determination– first clues

Regarding the analysis of sex determination, the starting point for this work was the hypothesis of genetic regulation, based on the strong correlations between genotype and sex found in *S. alburnoides* (reviewed in Alves *et al.*, 2001).

The main focus was on highly conserved genes that have been shown to take part in the process of sex determination throughout vertebrates. Although most have been extensively characterized in mammals, little is known about their function and regulatory interaction in most fish species. A global answer to the question of sex determination is apparently hard to find in teleost fish, due to the high diversity of mechanisms within the group and even between closely related species (Volff *et al.*, 2007). This constitutes an additional difficulty when considering sex determination in hybrid fish: parental species can often present different sex determination systems that are merged in the same individual. In the fish group, the evolutionary proximity of the genomes brought together by hybridisation is no guarantee of similarity of sex determination hierarchies and often results in alterations in the sex determination process and deviations in sex ratio distributions. As an example one could consider the closely related species of the genus *Xiphophorus*, which can exhibit XX/XY systems with occasional autosomal influence, ZZ/ZW and even polyfactorial systems (Vollf & Schartl, 2001). Crossing experiments involving two *Xiphophorus* species (*X. maculatus* and *X. hellerii*)

harboring different sex determination mechanisms were shown to result in a strong bias towards the male sex in the F1 hybrids (Woolcock *et al.*, 2006).

There is a general scarcity of information regarding the genetic basis of sex determination in hybrid fish, and the *S. alburnoides* complex exhibits an outstanding set of features that make it a desirable model to address those questions. Due to the particular interest on what deviations could account for an all male lineage of *S. alburnoides*, most of the genes were selected for their reported implication in male development in other vertebrate species, although an initial overlook on the alternative "female" pathway was attempted by following *dax1* and *figla* expression (Iyer & McCabe, 2004; Onichtchouk *et al.*, 2003). The behavior of these genes was analysed at three distinct levels. While their expression during development at presumptive gonad locations could imply participation in the sex determination process (and at later stages, the involvement in gonad differentiation mechanisms), their potential roles in adult gonad maintenance could be accessed by expression in particular cell types.

In *S. alburnoides*, gonad structure maturation and regression occurs cyclically during the annual breeding season (Ribeiro *et al.*, 2003), in which morphological identification of sexes and sex ratio calculations have always been performed. In other fish species, individual genes within the current set have been implicated in the functionality of specific cell types, in the safeguarding of adult gonad integrity and, directly or indirectly, in germ cell maturation (Rodríguez-Marí *et al.*, 2005; Nakamoto *et al.*, 2007; Kobayashi *et al.*, 2008; Klüver *et al.*, 2008). In this context it was interesting to explore whether the normal patterns of expression of such genes would also be maintained in the hybrids or if, conversely, their alteration could have some impact in gonad differentiation patterns. On the other hand, such deviations could strongly affect gonad functionality and even the success in the production and maturation of viable gametes, which would not agree with the full gonad functionality reported for most *S. alburnoides* forms (reviewed in Alves *et al.*, 2001). No differences were found in cellular location and in the expression patterns of individual genes in gonads of the parental species *S. pyrenaicus* and in the hybrid *S. alburnoides* (Chapters 2 and 3). The analysis did not involve more premature stages of gonad maturation (e.g. outside reproductive season), but the reproducibility of the location of individual gene expression and its correlation with specific cell types, even in an heterogeneous context of germ cell development, implies that normal patterns of gene expression are being resumed even upon hybridisation and that the normal process of cyclical differentiation and maintenance of gonad identity is taking place in hybrids as in parental species.

The participation of *amh*, *dmrt1*, *wt1* and *dax1* during *S. alburnoides* development was also assessed for the first time in the present work (Chapters 2 and 3). The first three genes were shown to be present at early stages of male development, in locations compatible with the developing gonads,

suggesting a possible participation in the sex determination process. Conversely, no *dax1* expression was found in early stage *S. alburnoides* male embryos, making its participation in male gonad fate determination highly improbable. It was possible to assess that the chosen candidates were in place, but the real functional contribution of these genes in shaping the sex determination process in *S. alburnoides* will still have to be determined. Such confirmation is impaired by an important difficulty that was not possible to overcome during the elaboration of this thesis: the absence of an effective and unambiguous method of identifying females during development. Several attempts have been made to overcome this problem, by finding an expression marker that could be used to identify sexes during development, but none proved to be effective. Genes that have reportedly been used as early female markers in vertebrates in general (Soyal *et al.*, 2000) and also in fish (Kanamori, 2000), such as factor in the germ line alpha, *figla* were isolated in *S. pyreanicus* and *S. alburnoides*, surprisingly showing no sexual dimorphic expression patterns in the adult and thus offering little confidence in their use as early female markers (it should be pointed out that there is no possibility of confirming sex of embryos by other means, so the genes of choice should offer a high confidence of exclusive female expression) (Chapter 3). Equally unexpected expression patterns were obtained for the aromatase gene *cyp19a,* responsible for the conversion of androgens to estrogens, by catalyzing the formation of estradiol 17-ß from testosterone (Scholz *et al.*, 2000), and reported as a female expression marker (Suzuki et al, 2004; Jørgensen *et al.*, 2008). A small partial fragment of the gene was obtained in *S. alburnoides* and *S. pyrenaicus*. The obtained product shows a higher homology to gonad-type aromatase (as opposed to brain-type aromatase, which would expectedly be expressed in both sexes) and the transcript was isolated both in males and females of *S. alburnoides* (Appendix 2), impairing its use as a specific sex marker. The possible reasons accounting for this similarity between male and female expression profiles will have to be further explored. The possibility of maintenance of a certain plasticity in shifting the differentiation of gonad structures into one or the other sex should always be considered, as it has been shown to occur in fish at given developmental stages (Uchida *et al.*, 2002) or at certain phases of the adult lifecycle (Liarte *et al.*, 2007). In the case of aromatase *cyp19a,* a longer gene product should be obtained to further confirm the orthology and potential functionality of the isolate obtained for both sexes. Although this gene has been classically correlated to the female sex, it should also be considered that *cyp19a* expression has been shown to occur in both adult ovaries and testis in some fish species, so more information should be gathered before making further assumptions based on the patterns observed in *S. alburnoides*. The possibility of transcript presence during development of both sexes and the determination of one or the other pathway through dosage effects as shown in the Japanese flounder, *Paralichthys olivaceus* (Kitano *et al.*, 1999) and zebrafish (Trant *et al.*, 2001) for *cyp19a,*

might constitute an additional disadvantage in non-quantitative approaches to its use as developmental sex marker in the *Squalius* species.

Alternative hypotheses should then be considered, namely ones with a more morphological basis. In the medaka *O. latipes*, the first morphological signal of gonad differentiation between sexes corresponds to the differences in germ cell proliferation: at stage 38, one or two days before hatching, the activity of germ cell division in female embryos increases, while male germ cells arrest in mitosis (Matsuda, 2005). Differences in proliferative state were also observed in other fish such as the nile tilapia *O. niloticus*, in which the first meiosis in the ovary occurs around 28 days after fertilization, while they are only initiated in the testes around 52-53 days after fertilization (Berishvili *et al.*, 2006). A more detailed morphological classification of the various stages of embryonic and gonad development would be necessary in *S. alburnoides* to assess whether similar disparities could be used to distinguish sexes at larval stages. In addition, markers for meiotic cells such as *scp3* (Lammers *et al.*, 1994) could be tested in order to investigate possible sex-specific differences in germ cell proliferative stages.

The first basis for a more extensive study of the genetic sex determination process in *S. alburnoides* has been launched and the first answers are only now starting to be obtained. However, the results obtained here have yet another level of impact, by making an additional contribution to the global knowledge of sex determination processes in fish. Even though hybridisation has a very high preponderance among teleosts (Le Comber & Smith, 2004), gene based approaches on sex determination have seldom been conducted in hybrid fish. The expression patterns observed for sex determination genes *amh*, *wt1* and *dax1* in *S. alburnoides* are apparently in accordance to what has been described for other teleost species, implying that there is in fact a good conservation of players and functions that may "resist" genome remodeling events such as hybridisation. But some interesting new features have also emerged. Unlike what has been described for many fish species (Guan *et al.*, 2000; Marchand *et al.*, 2000; Kobayashi *et al.*, 2008), expression of *dmrt1* in gonads of both sexes was shown in the present work, in *S. alburnoides* (Chapter 3). In the model cyprinid zebrafish, *dmrt1* also shows non-dimorphic expression according to sexes, although with a different cell-specific commitment (Guo *et al.*, 2005). Contrasting to expression in developing germ cells in both sexes of zebrafish, *dmrt1* expression in *S. alburnoides* was apparently correlated with supporting somatic cells in males and with later stage oocytes in females, thus suggesting that this gene may also have other unreported roles. It is possible that in addition to the direct commitment to germ cells (specially in males), *dmrt1* might also play a relevant role in the modulation of oocyte progression and additionally participate in the maintenance of testis structure and integrity, and indirect regulation of germ cell differentiation, as reported for other genes expressed in the Sertoli

cell lineage (Rodríguez-Marí *et al.*, 2005). These findings acquire a special interest because for many species *dmrt1* has been regarded as much more than another sex determination gene (reviewed in Ferguson-Smith, 2007). Despite the uncommon pattern of female *dmrt1* expression described in *S. alburnoides,* during development its classically reported participation in male gonad differentiation might also be conserved in the Iberian hybrids. *Dmrt1* is expressed in male embryos of *S. alburnoides* at developmental stages that might be relevant for sex determination. In the future it would be interesting to explore whether *dmrt1* would have a more upstream commitment in the male lineage, participating in the triggering of the male gonad "fate". Conversely, if as proposed for zebrafish (Guo *et al.*, 2005) it does participate in male and female gonad differentiation, how would its expression be modulated during development and it would correlate to the patterns observed in the adult.

Therefore, rather than a definitive answer, the present work has provided the first clues and the initial raw working basis that can be expanded in the search for a new overlook on gene-mediated sex determination in *S. alburnoides*. Hopefully it will constitute an additional step towards the elucidation of the complex and diverse pathways of sex determination in fish.

6.3. Non additive gene expression in *S. alburnoides*- new findings in a vertebrate context

6.3.1. 1+1+1=3?

Differential genome contribution to overall expression in triploid hybrids

Another main objective of the present work was to determine how the dynamics of genome-specific gene expression would be altered as a result of hybridisation and polyploidy, or in other words, whether all genome copies present in the hybrid would have an equal contribution to overall expression in the organism. Although the interest was to assess this for all hybrids in general, the study of higher ploidy forms offered an additional appeal, as it opened the possibility of assessing whether heterogeneous genome context or chromosome number rise would have a greater impact on the gene expression patterns. The theoretical expectation would be that an increase in copy number of all chromosomes would virtually affect all genes to the same extent, resulting in a uniform increase in gene expression. Focusing the study of gene expression patterns in triploid PAA hybrids of southern populations of *S. alburnoides* (the most common form in these locations), it was shown that genomes are contributing differentially to the expression of the analysed gene set, and that a mechanism of gene copy silencing is participating in the regulation of gene expression in *S.*

alburnoides. Interestingly, expression patterns were shown to vary according to gene and organ, suggesting a considerable plasticity in the process and rejecting the hypothesis of whole haplome silencing (Chapter 4).

These observations were in accordance to a number of gene regulation phenomena already reported in polyploid plants, showing similar patterns of differential expression according to organs (Adams et al, 2003, 2004) and non-addictiveness of expression following gene copy rise (Auger *et al.*, 2005; Wang *et al.*, 2006). In plants, these features have been accounted as important motors for polyploid species maintenance and evolution. The creation of redundancy by itself offers limited advantages, and its true evolutionary value mainly depends on the capacity of overcoming immediate effects on individual phenotype and fitness, maintaining the merged functionality of duplicates long enough for them to be "presented" for adaptation and evolutionary reshaping. Such effects have been described here for the first time in a polyploid vertebrate, but in evolutionary terms one could say that the implications are somehow similar to the ones described for other polyploid organisms. In vertebrates, as in plants (Adams *et al.*, 2003) gene silencing might be an important response to polyploidy, contributing to overcome the genetic instability brought on by odd genome numbers and increased gene copies, to generate phenotypic variability and maintain hybrid compatibility.

In functional terms, the gene copy silencing patterns described here for *S. alburnoides* implied that, unlike the initial reasoning for vertebrate polyploid taxa, the need for regulating the balance of gene products might also be relevant as a response to ploidy rise in animals (as proposed in Mable, 2007). Thus, the response to increased gene dosage by gene copy silencing observed in the *S. alburnoides* complex, implies that it can be a real possibility in polyploid taxa and that in fact it might be one of the factors contributing to the success of polyploids among lower vertebrates.

In the particular case of *S. alburnoides*, it was not possible to propose a mechanism to account for gene copy silencing, but the general mechanisms of epigenetic remodeling should offer some clues. Gene silencing could result from changes in base methylation or acetylation states, or higher order positional effects resulting from alterations in chromatin structure (Liu & Wendel, 2003). These processes can be directionally regulated and affect specific groups of genes, depending on chromosomal context, dosage requirements and parent of origin effects (Comai, 2000). In *S. alburnoides*, the dynamics of genome inheritance and the observed patterns of gene expression apparently rule out parental imprinting and imply a more random basis for the epigenetic processes that might potentially be occurring in the complex.

Although epigenetic remodeling might persist through evolutionary time, other more permanent changes have also been proposed to account for gene silencing, namely dispersed repetitive

element mediated rearrangements (Comai *et al.*, 2003). In *S. alburnoides*, the activity of transposable elements has been accounted as a possibility to justify the variation in the location of the major ribosomal genes within individuals, although this interaction has not been directly demonstrated (Gromicho, 2006). Besides more deterministic justifications, it should be considered that gene silencing observations could also be a side effect of mechanistic processes occurring in the cell or differential affinities in the interaction with protein complexes, and thus result from functionally neutral events (Comai, 2000). Also, it has been proposed that individual genes might be differentially affected by the parallel or superimposed action of a number of the referred processes. Epigenetic changes resulting from modifications in DNA methylation, histone modifications and chromatin structure alterations have been reported as possible causes underlying organ specific gene silencing in plants. It has been proposed that these changes are not mutually exclusive, and that different control levels might be involved in the expression regulation of individual genes (Adams et al, 2003). Therefore, and due to the plasticity of gene silencing patterns exhibited by *S. alburnoides*, it is possible that various of this factors might be contributing to the final outcome and hopefully in the future the mechanistic basis of this complex regulation system will start to be unraveled.

6.3.2. Heterogeneous gene expression patterns in different populations of the complex-evolutionary implications

At this point, the description of a genome regulation process had apparently been tackled in *S. alburnoides*, or at least the first hints of how it would be occurring in its most common form had been gathered, but a question remained: should gene expression patterns, that in plants have been accounted as one of the major forces shaping polyploid evolution (Adams & Wendel, 2005; Chen, 2007), be correlated to the evolutionary progression of the various populations of *S. alburnoides*? Even though global features are maintained, the history of the complex is not the same in different geographical locations (Alves *et al.*, 1997; Cunha *et al.*, 2004; Sousa-Santos *et al.*, 2007) and factors contributing to its evolutionary potential are also very distinct in the various populations (Pala & Coelho, 2005; Crespo-López *et al.*, 2007; Cunha *et al.*, 2008). Representative samples of this "diversity of fates" were probed and surprising results were obtained: a heterogeneous pattern of gene expression dynamics was found, correlating to geographical location. While in southern *S. alburnoides* polyploids a preferential silencing of P genome alleles was observed, in northern populations both C and A genome alleles were shown to be involved in overall expression, with no apparent bias in the contribution of one or the other genome. The exact participation of each

genome at the quantitative level is yet to be assessed in northern samples, but the implications of qualitative disparity between northern and southern allelic contribution should be considered for further discussion.

Several hypotheses have been postulated to correlate the observed patterns with the evolutionary history of the different lineages within *S. alburnoides,* and extensively discussed in Chapter 5. Viability and adaptation directed hypotheses have been proposed, namely considering the advantage potentially offered by the heterologous expression of alleles and the promotion of heterosis (Comai, 2005) in genetically impoverished populations. Also, the differential genomic context was also pointed out as a possibility to account for the observed differences. Most of the deviations in gene expression and epigenetic remodeling events in polyploid hybrids occur as a result from the merging of independent gene hierarchies and the need for maintaining a correct gene expression balance (Birchler & Veitia, 2007). It could be possible that C and A genomes could behave differently when brought together in the same nucleus, comparatively to P and A genomes. An interesting perspective comes from the study of interspecific variation of microRNAs and their impact on gene expression in allopolyploids (Ha *et al.*, 2008). Although regulatory and target recognition mechanisms of miRNAs show an overall conservation between species, variation in terms of timing, location and specific targets exists among even closely related species (Niwa & Slack, 2007). In allopolyploids, the gathering of two genome contexts might impact on the regulatory balance of miRNAs and their targets, resulting in differential patterns of gene expression. Some miRNAs may play a role in chromatin modifications and gene transcription (Bao *et al.*, 2004) and interestingly it has been shown that over 50% miRNA targets are among the genes that have been shown to have non-additive expression (Wang *et al.*, 2006) in plant allopolyploids (Ha et al, 2008). Although it is not possible to say whether such process should also be relevant in the case of *S. alburnoides*, the example aims at further demonstrating the diversity of factors that can modulate gene expression in allopolyploids and how they might directly depend on the particular genome hierarchies that are bought together (Riddle & Birchler, 2003). In evolutionary terms the work performed on *S. alburnoides* further confirmed that in vertebrates, as in plants, there could be much more than one solution for the allopolyploid enigma. Duplicate genes can be maintained through dissimilar regulatory means, and the escape from hybrid and odd genome constrictions can apparently be as effectively accomplished by “mimicking” a diploid context through allele-specific silencing or reaching for symmetry through the evolution by tetraploidisation. In a simplification, the options for Evolution are multiple and much depends on context. However, the study of polyploids offers some complex shades, much is yet to be uncovered and hopefully, by bringing the redundant

gene regulation world to a vertebrate context, new evolutionary questions have been brought to life.

6.4. References

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U S A. 100, 4649-4654.

Adams, K.L., Percifield, R., Wendel, J.F., 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168, 2217-2226.

Adams, K.L., Wendel, J.F., 2005. Polyploidy and genome evolution in plants. Curr Opin Plant Biol. 8, 135-141.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., Dowling, T.E., 1997. Maternal ancestry of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae) as determined by analysis of cytochrome b sequences. Evolution 51, 1584 – 1592.

Alves, M.J., Coelho, M.M., Collares-Pereira, M.J., 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111, 375 – 385.

Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H., Birchler, J.A., 2005. Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics 169, 389-397.

Bao, N., Lye, K.W., Barton, M.K., 2004. MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome, Dev. Cell. 7, 653–662.

Berishvili, G., D'Cotta, H., Baroiller, J.F., Segner, H., Reinecke, M., 2006. Differential expression of IGF-I mRNA and peptide in the male and female gonad during early development of a bony fish, the tilapia *Oreochromis niloticus*. Gen Comp Endocrinol. 146, 204-210.

Birchler, J.A., Veitia, R.A., 2007. The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell. 19, 395-402.

Chen, Z.J., 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids, Annu Rev Plant Biol. 58, 377–406.

Comai, L., 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol. 43, 387–399.

Comai, L., Madlung, A., Josefsson, C., Tyagi, A., 2003. Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? Philos Trans R Soc Lond B Biol Sci. 358, 1149-1155.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6, 836-846.

Crespo-López, M.E., Duarte, T., Dowling, T., Coelho, M.M., 2006. Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. Zoology (Jena) 109, 277-286.

Crespo-López, M.E., Pala, I., Duarte, T., Dowling, T.E., Coelho, M.M., 2007. Genetic structure of the diploid–polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. J. Fish Biol. 71, 423-436.

Cunha, C., Coelho, M. M., Carmona, J. A., Doadrio, I. 2004 Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridisation events. Mol Ecol. 13, 2807–2817.

Cunha, C., Doadrio, I., Coelho, M.M., 2008. Speciation towards tetraploidization after intermediate processes of non-sexual reproduction. Philos Trans R Soc Lond B Biol Sci. 363, 2921-2929.

Ferguson-Smith, M., 2007. The evolution of sex chromosomes and sex determination in vertebrates and the key role of DMRT1. Sex Dev. 1, 2-11.

Gromicho, M., 2006. Genome polyploidisation in the Iberian fish complex *Squalius alburnoides*. A case study of evolution-in-action through hybridisation and non-sexuality: a cytogenetic approach. PhD. Thesis, University of Lisbon, Lisbon, Portugal.

Guan, G., Kobayashi, T., Nagahama, Y., 2000. Sexually dimorphic expression of two types of DM (Doublesex/Mab-3)-domain genes in a teleost fish, the Tilapia (*Oreochromis niloticus*).Biochem Biophys Res Commun. 272, 662-666.

Guo, M., Rupe, M.A., Zinselmeier, C., Habben, J., Bowen, B.A., Smith, O.S., 2004. Allelic variation of gene expression in maize hybrids. Plant Cell 16, 1707-1716.

Guo, Y., Cheng, H., Huang, X., Gao, S., Yu, H., Zhou, R., 2005. Gene structure, multiple alternative splicing, and expression in gonads of zebrafish Dmrt1. Biochem Biophys Res Commun. 330, 950-957.

Ha, M., Pang, M., Agarwal, V., Chen, Z.J., 2008. Interspecies regulation of microRNAs and their targets. Biochim Biophys Acta. 1779, 735-742.

Iyer, A.K., McCabe, E.R., 2004. Molecular mechanisms of DAX1 action. Mol Genet Metab 83, 60-73.

Jørgensen, A., Morthorst, J.E., Andersen, O., Rasmussen, L.J., Bjerregaard, P., 2008. Expression profiles for six zebrafish genes during gonadal sex differentiation. Reprod Biol Endocrinol. 6, 25.

Kanamori, A., 2000. Systematic identification of genes expressed during early oogenesis in medaka. Mol Reprod Dev. 55, 31-36.

Kitano, T., Takamune, K., Kobayashi, T., Nagahama, Y., Abe, S.-I., 1999. Suppression of P450 aromatase gene expression in sex-reversed males produced by rearing genetically female larvae at a high water temperature during a period of sex differentiation in the Japanese flounder *(Paralichthys olivaceus*). J Mol Endocrinol. 23, 167–176.

Klüver, N., Herpin, A., Braasch, I., Drieβle, J., Schartl, M., 2008. Regulatory back-up circuit of medaka Wt1 co-orthologs ensures PGC maintenance. Dev Biol. [Epub ahead of print]

Kobayashi, T., Kajiura-Kobayashi, H., Guan, G., Nagahama, Y., 2008. Sexual dimorphic expression of DMRT1 and Sox9a during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*). Dev Dyn. 237, 297-306.

Lammers, J.H., Offenberg, H.H., van Aalderen, M., Vink, A.C., Dietrich, A.J., Heyting, C., 1994. The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes. Mol Cell Biol. 14, 1137-1146.

Le Comber, S.C., Smith, C., 2004. Polyploidy in fishes: patterns and processes. Biol. J. Linnean Soc. 82, 431-442.

Liarte, S., Chaves-Pozo, E., García-Alcazar, A., Mulero, V., Meseguer, J., García-Ayala, A., 2007. Testicular involution prior to sex change in gilthead seabream is characterized by a decrease in DMRT1 gene expression and by massive leukocyte infiltration. Reprod Biol Endocrinol. 5, 20.

Lim, K.Y., Soltis, D.E., Soltis, P.S., Tate, J., Matyasek, R., Srubarova, H., Kovarik, A., Pires, J.C., Xiong, Z., Leitch, A.R., 2008. Rapid chromosome evolution in recently formed polyploids in Tragopogon (*Asteraceae*). PLoS ONE. 3, e3353.

Liu, B., Wendel., J.F., 2003. Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogenet Evol, 29, 365-379.

Mable, B.K., 2007. Sex in the postgenomic era. Trends Ecol Evol. 22, 559-561.

Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J., Bernot, A., Laudet, V., Guiguen, Y., 2000. DMRT1 expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. Biochim Biophys Acta 1493, 180-187.

Matsuda, M., 2005. Sex determination in the teleost medaka, *Oryzias latipes*. Annu Rev Genet. 39, 293-307.

Nakamoto, M., Wang, D.S., Suzuki, A., Matsuda, M., Nagahama, Y., Shibata, N., 2007. Dax1 suppresses P450arom expression in medaka ovarian follicles. Mol Reprod Dev. 74, 1239-1246.

Niwa, R., Slack, F.J., 2007. The evolution of animal microRNA function. Curr Opin Genet Dev. 17, 145-150.

Onichtchouk, D; Aduroja, K; Belting, HG; Gnugge, L; Driever, W., 2003. Transgene Driving Gfp Expression From the Promoter of the Zona Pellucida Gene Zpc Is Expressed in Oocytes and Provides Marker for Gonad Differentiation in an Early Zebrafish. Dev Dyn. 228, 393–404.

Pala, I., Coelho, M. M., 2005. Contrasting views over a hybrid complex: between speciation and evolutionary "dead-end". Gene 347, 283–294.

Pontes, O., Neves, N., Silva, M., Lewis, M.S., Madlung, A., Comai, L., Viegas, W., Pikaard, C.S., 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. Proc Natl Acad Sci USA. 101, 18240-18245.

Ribeiro, F., I. G. Cowx, P. Tiago, A. F. Filipe, L. Moreira da Costa M. J. Collares-Pereira, 2003. Growth and reproductive traits of diploid and triploid forms of the *Squalius alburnoides* cyprinid complex in a tributary of the Guadiana River, Portugal. Arch. Hydrobiol. 156, 471–484.

Riddle, N.C., Birchler, J.A., 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Genet. 19, 597-600.

Robalo, J.I., Sousa Santos, C., Levy, A., Almada, V.C., 2006. Molecular insights on the taxonomic position of the paternal ancestor of the *Squalius alburnoides* hybridogenetic complex. Mol Phylogenet Evol. 39, 276-281.

Rodríguez-Marí, A., Yan, Y.L., Bremiller, R.A., Wilson, C., Cañestro, C., Postlethwait, J.H., 2005. Characterization and expression pattern of zebrafish Anti-Müllerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. Gene Expr Patterns. 5, 655-667.

Scholz, S., Gutzeit, H.O., 2000. 17-alpha-ethinylestradiol affects reproduction, sexual differentiation and aromatase gene expression of the medaka (*Oryzias latipes*). Aquatic Toxicol. 50, 363–373.

Sousa-Santos, C., Collares-Pereira, M.J., Almada, V., 2007. Reading the history of a hybrid fish complex from its molecular record. Mol Phylogenet Evol. 45, 981-996.

Soyal, S.M., Amleh, A., Dean, J., 2000. FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. Development 127, 4645-4654.

Stupar, R.M., Hermanson, P.J., Springer, N.M., 2007. Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. Plant Physiol. 145, 411-425

Suzuki, A., Tanaka, M., Shibata , N,, Nagahama, Y., 2004. Expression of aromatase mRNA and effects of aromatase inhibitor during ovarian development in the medaka, *Oryzias latipes.* J Exp Zoolog A Comp Exp Biol. 301, 266 –273.

Trant, J.M., Gavasso, S., Ackers, J., Chung, B.C., Place, A.R., 2001.Developmental expression of cytochrome P450 aromatase genes (CYP19a and CYP19b) in zebrafish fry (Danio rerio). J Exp Zool. 290, 475-483.

Uchida, D., Yamashita, M., Kitano, T., Iguchi, T., 2002. Oocyte apoptosis during the transition from ovary-like tissue to testes during sex differentiation of juvenile zebrafish. J Exp Biol. 205, 711–718.

Udall, J.A., Swanson, J.M., Nettleton, D., Percifield, R.J., Wendel, J.F., 2006. A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics 173, 1823-1827.

Volff, J.N., Schartl, M., 2001. Variability of genetic sex determination in poeciliid fishes. Genetica 111, 101–110.

Volff, J.-N., Nanda, I., Schmid, M., Schartl, M., 2007. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev 1, 85-99.

Wang. J., Tian, L., Lee, H.S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C. , Doerge, R.W., Comai, L, Chen, Z.J., 2006. Genomewide nonadditive gene regulation in Arabidopsis allotetraploids, Genetics 172, 507–517.

Woolcock, B., Kazianis, S., Lucito, R., Walter, R.B., Kallman, K.D., Morizot, D.C., Vielkind, J.R., 2006. Allele-specific marker generation and linkage mapping on the *Xiphophorus* sex chromosomes. Zebrafish 3, 23–37.

CHAPTER 7 | Concluding remarks

7. CONCLUDING REMARKS

The present work focused on the functional impact of hybridisation and polyploidy on gene expression patterns, and their evolutionary consequences. A complex challenge was faced when using the *S. alburnoides* as a starting point to address these questions, a system with an overwhelming potential for the study of the evolution of polyploids, but for which virtually no molecular features regarding genome organization and gene regulation had ever been described. The obvious outcome would be that in the attempt to address some of these questions an increasing number of new ones would arise. This last chapter summarizes the main achievements of this thesis, in relation to the initial objectives and the major questions and future research lines that have been generated from the present work.

7.1. The bigger picture: what have we learnt from the inspection of specific processes and global patterns of gene expression?

In an integrative view, the present work constituted the first step towards an understanding of the complexity of genome regulation and gene expression in a vertebrate allopolyploid context. It also constitutes an initial approach to the genetic basis of an important feature in the dynamics of the *S. alburnoides* complex: sex determination and the establishment of sex ratios.
In summary, the main achievements of this work were:

1. Through the isolation and subsequent characterization of the structural features and expression patterns of six genes shown to participate in vertebrate sex determination (*amh*, *dmrt1*, *wt1*, *dax1, figla* and *vasa*), a molecular basis for the study of this process was created in the hybrid system of *S. alburnoides.* A participation of conserved genes in vertebrate sexual development was observed during *S. alburnoides* development. Early expression of *amh*, *dmrt1* and *wt1* was observed in male embryos of *S. alburnoides*, in locations compatible with the developing gonad, thus suggesting that also in the hybrid complex, these genes might have a relevant role in the establishment of male gonad structures.

2. For the first time in the study of the complex, the possibility of a morphological characterization of the adult gonads was created, for individuals of *S. pyrenaicus* and *S. alburnoides*. The organizational structure of the male gonad was further inspected through immunoprecipitation assays in an attempt to determine the presumptive location of specific cell types. A better understanding of the

cell type distribution and the composition of *Squalius* gonads was obtained through the comparison with gonad morphologies of other teleosts, allowing for the association of particular gene expression patterns to specific cell types.

3. For all genes, the adult expression patterns were analysed in *S. alburnoides* and *S. pyrenaicus* and specific locations of expression were determined both in male and female gonads. The presumptive contribution of these genes to gonad integrity maintenance and functionality was apparently conserved both in the bisexual species and in *S. alburnoides*, which correlates well with the reported fertility of the various forms that compose the complex. Therefore, it could be established that no abrupt shifts on gene expression dynamics would have occurred in *S. alburnoides* as a result of the hybridisation process and that individual gene contribution, at least in terms of location, would globally remain unaltered in the hybrids.

4. A gene expression regulation mechanism involving gene copy silencing was formally described here, for the first time, in a polyploid vertebrate. Through the analysis of a set of seven housekeeping and tissue specific genes, genome-specific allele silencing was observed in triploid PAA individuals of *S. alburnoides*. Transcript levels were apparently reduced to the diploid state in triploids, suggesting that the observed allele silencing could occur as a response to increased gene dosage. It was shown that silencing was not due to whole haplome inactivation and that expression patterns vary according to gene and organ. The report of a dosage compensation mechanism in such an emblematic system as *S. alburnoides* opened other hypotheses in the study of polyploid vertebrate genomes, namely by introducing gene dosage balance, the occurrence of regulatory phenomena of gene copy inactivation and non-additivity of gene expression as relevant features contributing to functionality and persistence of polyploids.

5. The unraveling of heterogenous patterns of genome-specific allelic expression in *S. alburnoides* was also an important achievement of this thesis. Patterns of gene expression and the specific contribution of each heteromorphic genome were assessed in northern and southern populations of *S. alburnoides*, for a total of seven genes, and consistent differences emerged correlating the patterns of allelic-specific contribution to geographical location. Several distinct evolutionary features have emerged from the characterization of the southern and northern populations of the complex, but the present work presented new additional evidence that could be highly relevant to the understanding of these differences. Apparently, various responses in terms of genome-specific gene copy expression can occur, resulting from particular genome contexts, even within the same

hybrid complex, as shown here in *S. alburnoides*. While in some populations, in particular conditions and according to specific genome combinations, silencing of gene copies of a particular genome might offer functional advantage, in others the same might not be true and different evolutionary pathways might be followed to allow for lineage progression.

7.2. Future perspectives

A number of new questions have arisen from this approach, and some perspectives that could be addressed in future research not only of the *S. alburnoides* complex but also of other polyploids are layed out in this section.

- Regarding sex determination, it would be necessary to determine more precisely the contribution of some of the isolated candidates, namely by establishing a comparison basis between females and males, during development. For this purpose, the establishment of molecular or morphologically based methods of sex identification should be highly advisable.

- Some interesting features have emerged from this thesis, namely for genes such as *dmrt1*, in which a participation in primordial germ cell fate was suggested by some experimental results obtained during this work. Thus, a more detailed analysis of the early expression of *dmrt1* in different male forms (hybrid and "nuclear non-hybrid") should be performed in the future. This search would be in line with an important question that remains to be answered in *S. alburnoides*: is there a gene that triggers a pathway to "mandatory" maleness in AA genotype individuals? *Dmrt1* might be a good starting point to address this question.

- If the search for answers regarding sex determination is to be taken to a new complexity level of gene interactions, then the addition of extra gene candidates should be considered. One possible candidate could be *sf1*, for which cooperation or antagonistic interactions with other members of the present list of genes isolated in *S. alburnoides* and *S. pyrenaicus* (namely *wt1* and *dax1*, and indirectly, the modulation of *amh* expression) have been reported in different groups.

- The reported association of gene expression to particular cell types in the adult gonads of *S. alburnoides* and *S. pyrenaicus* might also be additionally used as a basis for the study of gonad organization during the annual maturation period in the *Squalius* species. It might be interesting to follow the expression of some of these genes at several stages of this annual phase to assess how cellular hierarchies are being organized over time and whether these genes would have an

additional role in this process. Gonad maturation changes within the complex have seldom been characterized and never described at the histological level. At some points during the elaboration of this thesis, it would have been useful to have such detailed information to better interpret the patterns of expression of particular genes and their significance. It could be an interesting additional feature to explore in the study of *S. alburnoides.*

- A phenomenon of gene copy silencing in a polyploid context has been reported here for the first time in *S. alburnoides* but, within the scope of this thesis, it was not possible to determine the underlying mechanism by which silencing is effectively attained. The complexity of the process and the general lack of molecular data on *S. alburnoides* (namely access to gene promoters) increase the difficulty of this task. However, the unraveling of the mechanism which modulates gene copy expression in a polyploid context offers exciting perspectives that can by far surpass the difficulties, not only for the study of polyploid genomes but also to the understanding of what pathways might be disrupted in the case of aneuploidies. A possible starting point could be the comparative study of methylation patterns in the hybrids and in the bisexual maternal ancestor. For this, however, the suitable promoter regions would have to be identified and a more precise *a priori* knowledge of the actual gene expression interactions would have to be attained.

- Additionally, gene expression analysis could be extended to *S. alburnoides* in the area of sympatry with *S. aradensis*. The main aim would be to establish how an apparently more complex system of genomic interactions would react in terms of gene expression regulation and compare the observed patterns with the ones reported in the present work.

- Stepping slightly away from the *S. alburnoides* system, and considering the new perspectives that the present work brings to the study of fish polyploids, it would be interesting to assess whether the same patterns of gene expression regulation would be acting in other polyploid systems, with a completely distinct system of genetic exchanges and different evolutionary dynamics. An interesting perspective would also be the investigation of whether similar regulatory changes would be occurring in autopolyploid species or if, in vertebrates, they would mainly result from the interaction between different genomes.

APPENDIX 1

APPENDIX 1

Analysis of the expression dynamics of ribosomal genes- Isolation of the *5S*, *18S* and *28S* genes in *S. pyrenaicus* and *S. alburnoides*

As alterations to the expected expression dynamics and dosage compensation have been observed for several genes in *S. alburnoides*, it would be interesting to inspect whether ribosomal genes are influenced by similar regulation constrains. If genome specific or dosage related expression patterns were to be observed also for these genes, it is possible that a more global explanation to the regulation patterns of this hybrid complex could be proposed. The key role of these genes in translation makes them good candidates to start addressing this question.

A1.1. Materials and Methods

A1.1.2. Primer design and amplification

Primers for the *28S* and *5S* genes, already used in the study of the complex (Gromicho *et al.*, 2005) were tested, using cDNA and genomic DNA samples of *S. alburnoides* (AA genotype) and *S. pyrenaicus* (PP genotype) as templates. In the first case, organ samples of liver and gonad from *S. alburnoides and S. pyrenaicus* were used for RNA extraction and cDNA synthesis, while genomic DNA was obtained from fin clip samples.

New exon specific primers for the *28S* gene (Table A. I.) were designed, based on partial coding sequences of the zebrafish ortholog (AF398343), collected by database search. Primers for the *18S* gene, described in Filby & Tyler (2007) were additionally tested in cDNA samples of *S. pyrenaicus*, *S. alburnoides* and *S. carolitertii*.

Table A1. I. Primers based on the *28S* coding sequence of *D. rerio*.

Primer	Sequence
28S-F1	5'-ATGGCCACGCTGGAGRGCTGCC- 3'
28S-R1	5'- GCTGGTCGTCCTCSGGC - 3'
28S-R2	5'- TGCAGCACYTGYTSCAGAAG- 3'

Primers were tested according to the following PCR conditions: pre-heating at 96°C for 2 min 30 s, 35 cycles at 96°C for 30 s, 50°C-54°C for 30 s and 72°C for 1 min 15s and a final extension at 72°C for 10 min. In the cases for which amplification was successful, products were recovered from agarose gel and purified using E.Z.N.A. Gel extraction Kit (PeqLab), subsequently sequenced and analysed using Sequencher ver. 4.0 (Gene Codes Corporation, Inc.).

A1.2. Results

A1.2.1. *5S*

The amplification using cDNA templates of liver and gonad samples of the two species resulted in the production of a 111 bp PCR fragment (Figure A1.1). BLAST comparisons confirmed an overall identity with coding sequences of other vertebrate *5S* genes. However, the total identity between gene segments of *S. pyrenaicus* and *S. alburnoides* prevented any further polymorphism based analysis using the obtained region.

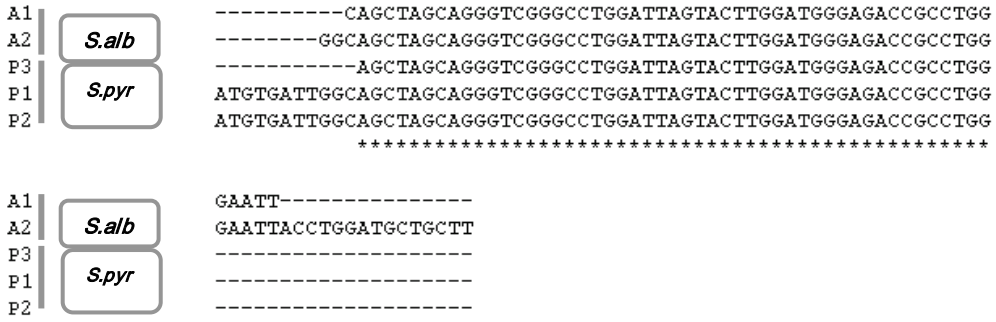


Figure A1.1. Alignment of partial *5S* sequences of *S. pyrenaicus* (*S.pyr*) and *S. alburnoides* (*S. alb*). (*) Identical residues in all sequences in the alignment; (:) Conserved substitutions; (.) Semi-conserved-substitution.

A1.2.2. *28S*

The primers designed specifically for cDNA, promoted the amplification of a 494bp product, which exhibited high identity to other teleost *28S* genes, through BLAST comparison. However, no polymorphisms were found between the A and P genomes that could be informative in terms of characterizing the expression differences between genomes (Figure A1.2). Although globally conserved, the region exhibits a number of polymorphic sites in other teleost species like *Tetraodon nigroviridis* (AJ270041) and *Oncorhynchus mykiss* (OMU34341) (Fig. A1.2). The present results might indicate that there is a high overall conservation of the coding sequence of the *28S* gene for the A

and P genomes, suggesting that this gene (or at least this region) might not be a good candidate to follow genome specific allele expression.

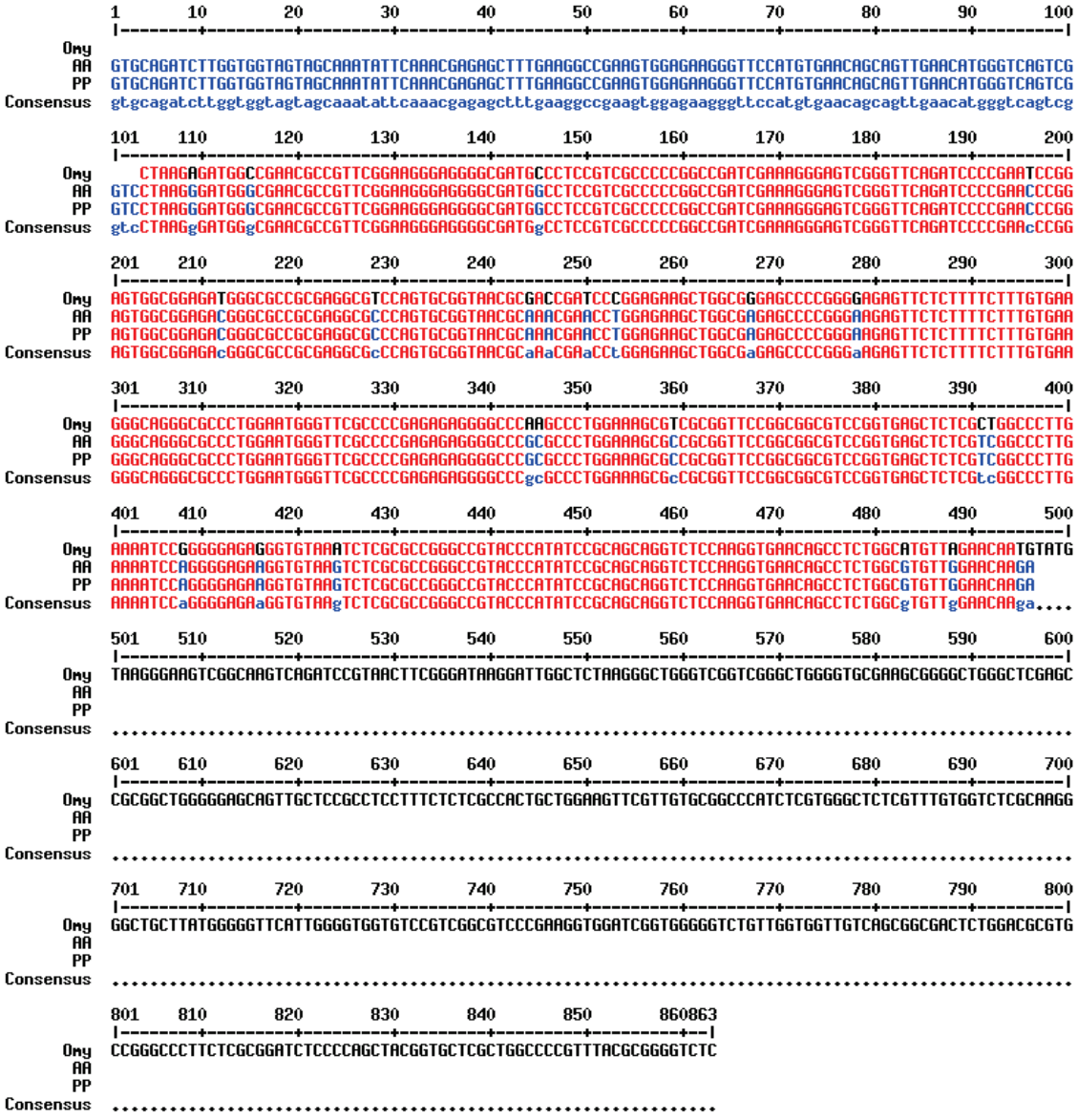


Figure A1.2. Alignment of *S. alburnoides* (AA), *S. pyrenaicus* (PP) partial coding sequence of the *28S* gene and the *O. mykiss* ortholog. Identical residues in all sequences in the alignment are indicated in red.

1.2.3. *18S*

The results obtained with the *18S* gene primers using *Squalius* samples as templates revealed an overall identity between the three genomes analysed: A, P and C (Fig. A1.3). BLAST comparisons confirmed an overall identity with coding sequences of other vertebrate *18S* genes. No polymorphic

sites were found between genomes, for the nine samples analysed, and this gene was at least temporarily abandoned as a candidate to follow genome specific allele expression.

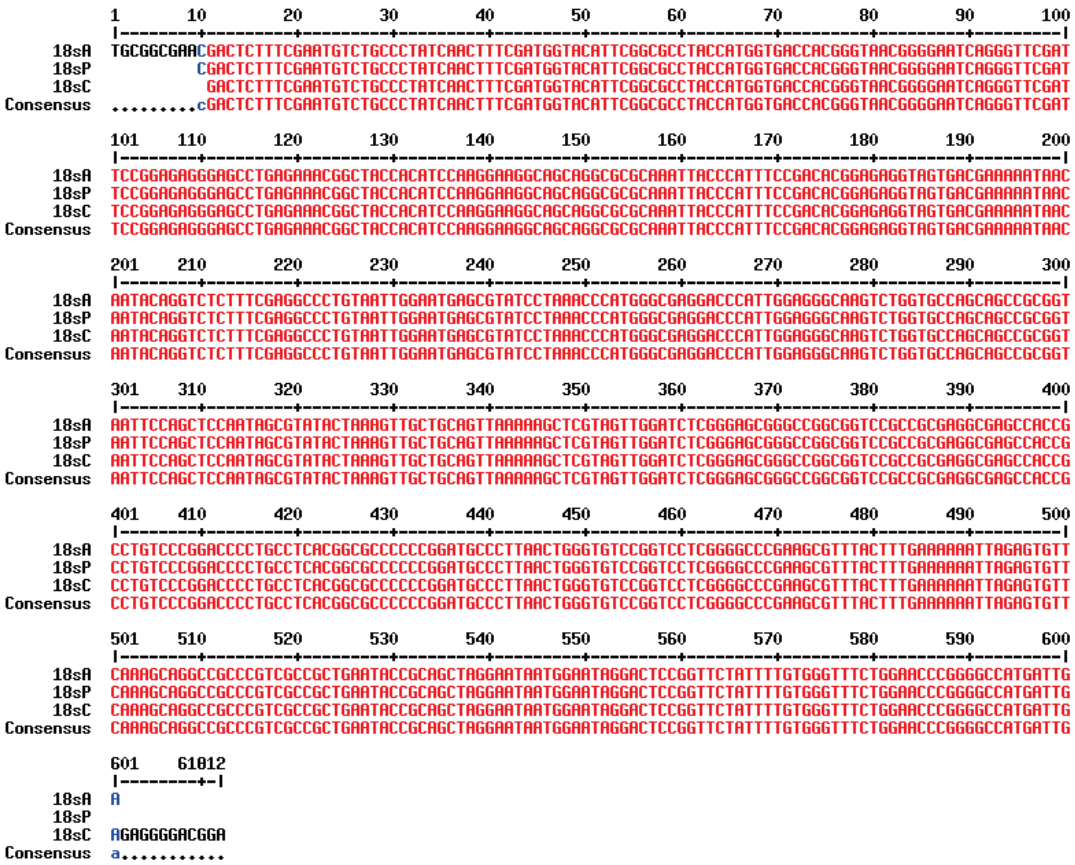


Figure A1.3. Alignment of partial *18S* sequences of *S. pyrenaicus* (*S.pyr*- P genome), *S. alburnoides* (*S. alb*- A genome) and *S. carolitertii* (*S. car*- C genome). Identical residues in all sequences in the alignment indicated in red.

A1.3. References

Filby, A.L., Tyler, C.R., 2007. Appropriate "housekeeping" genes for use in expression profiling the effects of environmental estrogens in fish. BMC Mol Biol 8, 8-10.

Gromicho, M., Ozouf-Costaz, C., Collares-Pereira, M.J., 2005. Lack of correspondence between CMA3-, Ag-positive signals and 28S rDNA loci in two Iberian minnows (Teleostei, Cyprinidae) evidenced by sequential banding. Cytogenet Genome Res. 109, 507-511.

APPENDIX 2

APPENDIX 2

In the search for a gene that could potentially be used as an early ovarian marker, the cytochrome P450 aromatase gene (*cyp19*) was isolated in *S. alburnoides* and *S. pyrenaicus*. It is responsible for the conversion of androgens to estrogens, and has been reported as a female expression marker in different groups (Suzuki *et al.*, 2004; Jørgensen *et al.,* 2008).

A2.1. Materials and Methods

A2.1.1. Primer design and amplification

Primers were designed based on the *Danio rerio* P450 aromatase (*cyp19a*) complete coding sequence (AF183906). The positioning of each primer on a separate exon was based on the aromatase transcript structure of zebrafish (ENSDART00000060605), thus ensuring a product size difference between genomic DNA and cDNA (Table A2.I).

Primer	Sequence	Location
Cyp9_F1	TTGTGCGTGTCTGGATCAAT	Exon 2
Cyp9_R1	TGAATATGATGCCCTGCTCA	Exon 3
Cyp9_F2	TCCGAATTCTTCTCCAAACC	Exon 8
Cyp9_R2	AATGTGCTTCCCCACACAC	Exon 9

Table A2.I. Primers based on the *cyp19a* coding sequence of *D. rerio*.

A total of six gonad samples of *S. alburnoides* (AA genotype) and *S. pyrenaicus* (PP genotype) were used as templates. RNA was extracted from all gonad samples and cDNA was synthesized using the RevertAid™ First Strand cDNA Synthesis Kit (Fermentas). Genomic DNA was additionally extracted from two *S. pyrenaicus* gonad samples and used as a control for size specificity of the obtained products.

A2.1.2. Isolation of partial sequences of the *cyp19* in *S. pyrenaicus* and *S. alburnoides*

The primers designed here resulted in positive amplification of products of approximately 400bp (F2/R1) and 1000bp (F1/R2). Female and male gonad samples had been tested and, surprisingly, amplification of the putative female expression marker was successful in samples of both sexes. A high sequence similarity was observed in the products obtained in males and females, both in *S. pyrenaicus* and *S. alburnoides*. BLAST comparisons confirmed an overall identity of the amplified

products with coding sequences of other vertebrate aromatase genes, thus confirming the isolation of a partial coding sequence of *cyp19* in *S. pyrenaicus* and *S. alburnoides*.

The amplification of aromatase in male gonad samples was unexpected and further analysis was performed to assess whether it could correspond to another non-gonad specific product. It was observed that the fragment amplified in *S. alburnoides* and *S. carolitertii* included the specific "substrate-binding-domain" of *cyp19* genes (Chiang *et al.*, 2001). Also, homologies with known aromatase types were also inspected in order to assess if there could be differences in the products amplified from male and female samples. The alignment of predicted amino acid sequences of the *Squalius cyp19* products with ovarian (*cyp19a*) and brain (*cyp19b*) type aromatase of zebrafish, confirmed the homology of all obtained products with the gonad-specific gene (Fig. A2.1).

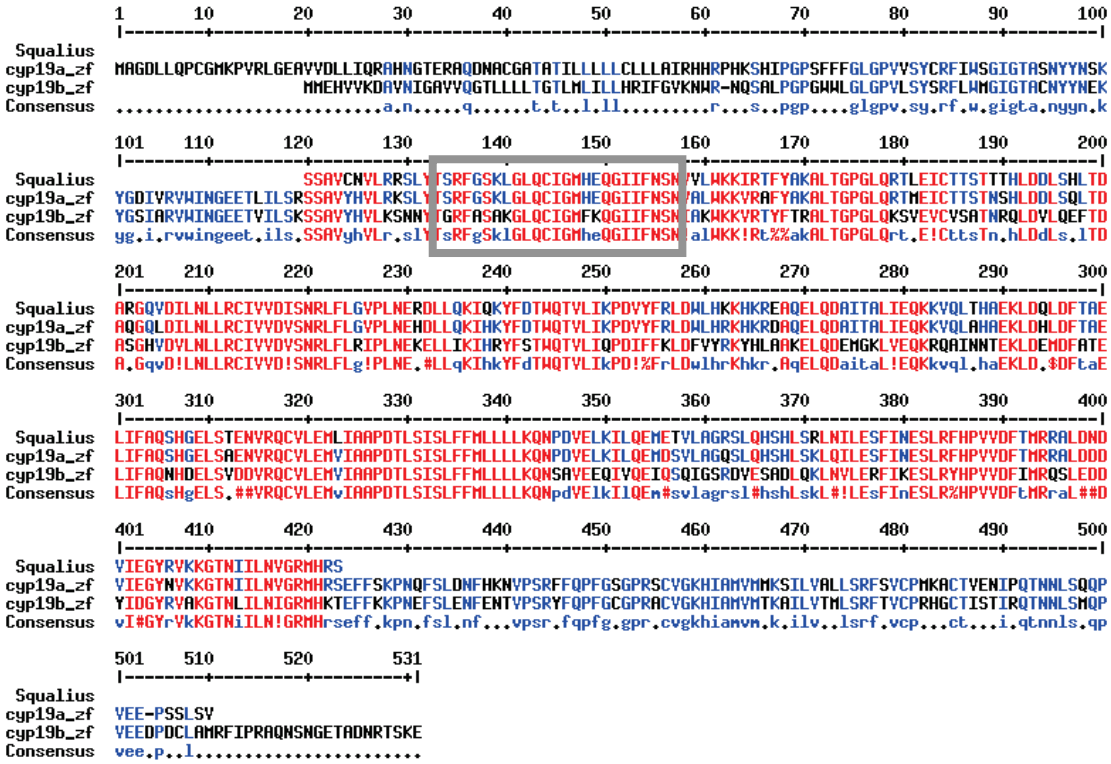


Figure A2.1. Protein alignment of a representative sequence of the Squalius *cyp19* with ovarian (cyp19_a) and brain (cyp19_b) type *aromatase* of the zebrafish *D. rerio*. Substrate-binding domain is highlighted in grey.

The apparent non-dimorphic expression of aromatase in the adult gonads of *Squalius* will have to be further investigated and the specific expression patterns of this gene should be assessed through *in situ* hybridisation. However, the results presented here prevented any further use of the aromatase gene *cyp19* as an ovarian-specific expression marker in *S. alburnoides* and *S. pyrenaicus*, at least until these features are further elucidated.

A2.2. References

Chiang, E.F., Yan, Y.L., Guiguen, Y., Postlethwait, J., Chung, Bc., 2001. Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. Mol Biol Evol. 18, 542-550.

Jørgensen, A., Morthorst, J.E., Andersen, O., Rasmussen, L.J., Bjerregaard, P., 2008. Expression profiles for six zebrafish genes during gonadal sex differentiation. Reprod Biol Endocrinol. 6, 25.

Suzuki, A., Tanaka, M., Shibata , N,, Nagahama, Y., 2004. Expression of aromatase mRNA and effects of aromatase inhibitor during ovarian development in the medaka, *Oryzias latipes.* J Exp Zoolog A Comp Exp Biol. 301, 266 –273.